Proceedings of the I·R·E

AUGUST 1941

VOLUME 29 NUMBER 8

Development of Sensitive Phototubes

Magnetic Recording in the Broadcast Field

Instruments for Measuring Radio Noise

Design of the Universal Winding

Unsymmetrical-Sideband Problems

Distant Field of Linear Radiators

Energy Equation in Electronics

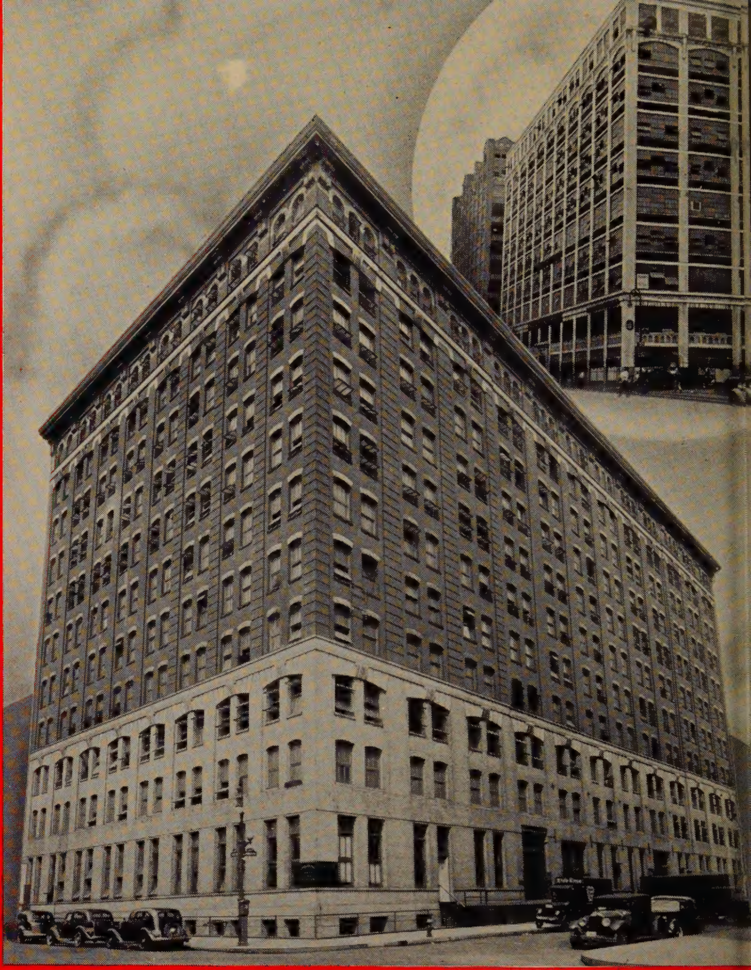
High-Frequency Transmission

Institute of Radio Engineers

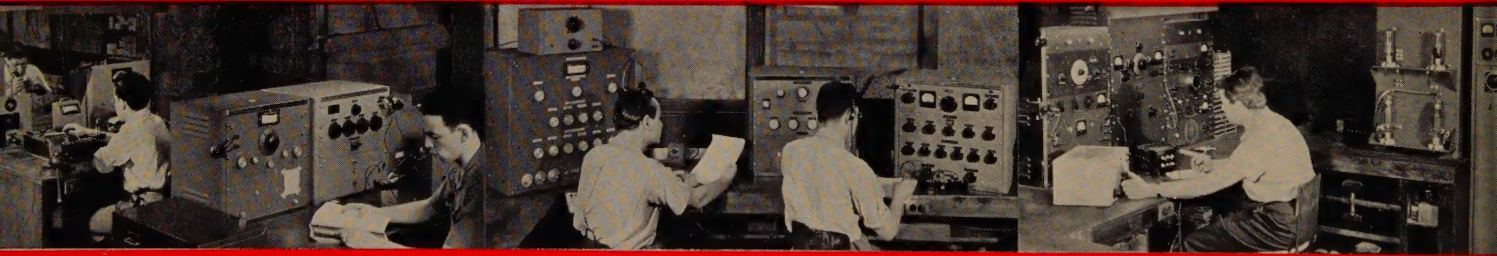
FOR BETTER DELIVERIES

ANOTHER UTC PLANT

Added facilities in another UTC plant now bring the size of UTC to thirty times that in 1933. This new plant assures a continuance of UTC's fine delivery record.



For Better Products . . . **Continuous Research**



As an organization founded on research and design ingenuity, the UTC research laboratories are unusually fine. A continuous research and development program is maintained to improve the quality and dependability of UTC products. In addition, customers' problems and their solutions are handled by a competent staff of application engineers.

UNITED TRANSFORMER CORP.

150 VARICK STREET ★ NEW YORK, N. Y.

EXPORT DIVISION: 100 VARICK STREET NEW YORK, N. Y.

Proceedings of the I·R·E

Published Monthly by
The Institute of Radio Engineers, Inc.

VOLUME 29 — *August, 1941* — NUMBER 8

A Review of the Development of Sensitive Phototubes... Alan M. Glover 413

Magnetic Recording and Some of Its Applications in the Broadcast Field.......... S. J. Begun 423

Progress in the Development of Instruments for Measuring Radio Noise.......... Charles M. Burrill 433

The Design of the Universal Winding.......... L. M. Hershey 442

The Solution of Unsymmetrical-Sideband Problems with the Aid of the Zero-Frequency Carrier.......... H. A. Wheeler 446

The Approximate Representation of the Distant Field of Linear Radiators.......... Ronold King 458

On the Energy Equation in Electronics at Ultra-High Frequencies.......... C. K. Jen 464

High-Frequency Radio Transmission conditions, July, 1941, with Predictions for October, 1941.......... 467

Institute News and Radio Notes.......... 469

- Editorial Procedure.......... 469
- Board of Directors.......... 469
- Sections.......... 470
- Membership.......... 473

Contributors.......... 473

Entered as second-class matter October 26, 1927, at the post office at Menasha, Wisconsin, under the Act of February 28, 1925, embodied in Paragraph 4, Section 538 of the Postal Laws and Regulations. Publication office, 450 Ahnaip Street, Menasha, Wisconsin. Editorial and advertising offices, 330 West 42nd St., New York, N. Y. Subscription, $10.00 per year; foreign, $11.00.

The Institute of Radio Engineers

INCORPORATED

New York Meeting—October 1, 1941

Rochester Fall Meeting, November 10, 11, and 12, 1941

SECTION MEETINGS

ATLANTA September 19	DETROIT September 19	PHILADELPHIA October 2
CHICAGO September 19	LOS ANGELES September 16	PITTSBURGH September 9
CLEVELAND September 25		WASHINGTON September 8

SECTIONS

ATLANTA—Chairman, A. W. Shropshire, Secretary, G. M. Howard, 856 St. Charles Ave., N. E., Atlanta, Ga.

BALTIMORE—Chairman, Ferdinand Hamburger, Jr.; Secretary, G. J. Gross, Pennsylvania Water and Power Co., 1611 Lexington Bldg., Baltimore, Md.

BOSTON—Chairman, J. M. Henry; Secretary, R. O. Oberg, Northeastern University, Boston, Mass.

BUENOS AIRES—Chairman, P. J. Noizeux; Secretary, L. C. Simpson, RCA Victor Argentina, Bartolome Mitre, 1961, Buenos Aires, Argentina.

BUFFALO-NIAGARA—Chairman, E. H. Roy; Secretary, Leroy Fiedler, 53 Rosedale Ave., Hamburg, N. Y.

CHICAGO—Chairman, G. I. Martin; Secretary, R. A. Kay, RCA Institutes, Inc., 1154 Merchandise Mart, Chicago, Ill.

CINCINNATI—Chairman, J. M. McDonald; Secretary, W. L. Schwesinger, Crosley Corp., Cincinnati, Ohio.

CLEVELAND—Chairman, Carl E. Smith; Secretary, W. G. Hutton, 7314 Dorothy Ave., Parma, Ohio.

CONNECTICUT VALLEY—Chairman, K. A. McLeod; Secretary, W. M. Smith, 250 Main St., Suffield, Conn.

DALLAS-FORT WORTH—Chairman, D. A. Peterson; Secretary, P. C. Barnes, WFAA-WBAP, Grapevine, Tex.

DETROIT—Chairman, M. Cottrell; Secretary, Paul Frincke, WJBK, 6559 Hamilton Ave., Detroit, Mich.

EMPORIUM—Chairman, R. K. Gessford; Secretary, C. W. Reash, Hygrade Sylvania Corp., Emporium, Pa.

INDIANAPOLIS—Chairman, A. N. Curtiss; Secretary, T. N. Rosser, P. R. Mallory & Co., 3029 E. Washington St., Indianapolis, Ind.

KANSAS CITY—Chairman, Harner Selvidge; Secretary, G. L. Taylor, Midland Radio School, Power and Light Bldg., Kansas City, Mo.

LOS ANGELES—Chairman, W. W. Lindsay, Jr.; Secretary, C. F. Wolcott, 1609 S. Western Ave., Los Angeles, Calif.

MONTREAL—Chairman, E. A. Laport; Secretary, W. A. Nichols, Canadian Broadcasting Corp., 1012 Keefer Bldg., Montreal, Que.

NEW ORLEANS—Chairman, G. H. Peirce; Secretary, D. W. Bowman, 8327 Sycamore St., New Orleans, La.

PHILADELPHIA—Chairman, C. C. Chambers; Secretary, R. L. Snyder, 103 Franklin Rd., Glassboro, N. J.

PITTSBURGH—Chairman, P. N. Bossart; Secretary, Ralph Overholt, Jr., 7211 Thomas Blvd., Pittsburgh, Pa.

PORTLAND—Chairman, E. R. Meissner; Secretary, L. M. Belleville, 361 N. E. 75th Ave., Portland, Ore.

ROCHESTER—Chairman, H. J. Klumb; Secretary, O. L. Angevine, Jr., Stromberg-Carlson Telephone Manufacturing Co., Rochester, N. Y.

SAN FRANCISCO—Chairman, L. J. Black; Secretary, Karl Spangenberg, Electrical Engineering Dept., Stanford University, Calif.

SEATTLE—Chairman, K. H. Ellerbeck; Secretary, R. J. Gleason, Pacific Alaska Airways, Inc., 3314 White Bldg., Seattle, Wash.

TORONTO—Chairman, R. H. Klingelhoeffer; Secretary, L. C. Simmonds, 301 King St., E., Toronto, Ont.

TWIN CITIES—Chairman, C. A. Culver; Secretary, N. B. Coil, 1664 Thomas St., St. Paul, Minn.

WASHINGTON—Chairman, M. H. Biser; Secretary, C. M. Hunt, WJSV, 817 Earle Bldg., Washington, D. C.

A Review of the Development of Sensitive Phototubes*

ALAN M. GLOVER†, ASSOCIATE, I.R.E.

Summary—*The development of phototubes is traced from early investigations of the nature of the photoelectric effect to the development of modern sensitized photosurfaces. The photoelectric properties of such surfaces are contrasted with those of pure metals. A discussion is given of the technical developments which have contributed to the importance of modern phototubes. A comprehensive bibliography covering both general and specific references on the subject of photoelectricity with emphasis on photoemissive surfaces is included.*

Historical Introduction

THE TERM "photoelectric cell" may be applied to three different devices which respond to the incidence of light or other electromagnetic radiation by the flow of an electric current. Of these three, the photoemissive cell or phototube is historically the most recent. In 1839, Becquerel[1,2] discovered that the illumination of one of two silver electrodes in an electrolytic or voltaic cell resulted in the production of an electromotive force which thus caused a current to pass through the cell. Such cells are now known as photovoltaic cells because the generation of the electromotive force was long thought to be the primary mechanism involved. The discovery that the resistance of selenium is altered upon illumination has been attributed to W. Smith who, in 1873,[3] was engaged in developing materials for use in the transatlantic cable. Light-sensitive materials in which a change in current is thus produced are known as photoconductive cells. In general, such substances are found to be semiconductors, although insulators exhibit a small photoconductive effect.

In 1887, Hertz[4] was engaged in an experiment which demonstrated the validity of Maxwell's prediction that energy should be radiated from a conductor in which a rapidly changing current flows. Hertz discovered that when a spark discharge took place in a circuit, a small spark was observed at a gap in a similar neighboring circuit. Incidentally, this experiment laid the basis for the development of radio. In the course of the experiment, Hertz found that the length of the spark induced in the auxiliary circuit was reduced if the light from the primary spark was not allowed to fall upon the second gap. The effect was traced to the ultraviolet light from the spark and was found to be primarily due to the illumination of the negative electrode. The investigations of Hallwachs[5] showed that negative electricity was released from a negatively charged zinc plate under the influence of ultraviolet light. This current was identified in 1899 by Lenard,[6] and by J. J. Thomson[7] with the electrons which constituted the current in the discharge of a Geissler tube. Elster and Geitel[8] showed in 1899 that certain metals such as potassium and sodium, when amalgamated with mercury, were sensitive even to visible light. Within a short time, they realized that a considerably enhanced and more stable emission current could be obtained if the negative element and a positive collector were enclosed in an evacuated bulb.[9,10] Hence, the phototube may be considered as one of the first of the family of electronic tubes. As early as 1892 Elster and Geitel[11,12] devised an instrument which incorporated a photosensitive surface for measuring the ultraviolet light from the sun.

The photoelectric current was also shown to be directly proportional to the intensity of the light producing it. However, the kinetic energy of the released electrons was found to be independent of the intensity of the light, but to be dependent upon the frequency of the light.[6] These facts, which were inconsistent with a wave theory of the nature of light, led Einstein[13] to postulate that light was corpuscular in nature and that each light corpuscle or "quantum" could give up its energy to an electron of the metal. He assumed that the energy of each quantum of light of a given frequency was proportional to the frequency and was equal to $h\nu$ where ν is the frequency and h is Planck's constant. The kinetic energy E of the electron released would then be equal to $h\nu$ less any energy lost by the electron in passing through the surface. This latter amount is usually expressed in terms of the potential which the electron must overcome in leaving the

* Decimal classification: 535.38. The original manuscript of this invited paper was received by the Institute, June 9, 1941.

† RCA Manufacturing Company, Inc., Harrison, N. J.

[1] E. Becquerel, "Recherches sur les effets de la radiation chimique de la lumiere solaire, au moyen des courants electriques," *Compt. Rend.*, vol. 9, pp. 144–149; July, 1839.

[2] E. Becquerel, "Memoire sur les effets electriques produits sous l'influence des rayons solaires," *Compt. Rend.*, vol. 9, pp. 561–567; November, 1839.

[3] S. Bidwell, "On the sensitiveness of selenium to light, and the development of a similar property in sulphur," *Phil. Mag.*, vol. 20, pp. 178–190; August, 1885.

[4] H. Hertz, "Ueber einen Einfluss des ultravioletten Lichtes auf die electrische Entladung," *Ann. der. Phys.*, vol. 31, no. 8 (B), pp. 983–1000; 1887.

[5] W. Hallwachs, "Ueber den Einfluss des Lichtes auf electrostatisch geladene Korper," *Ann. der Phys.*, vol. 33, no. 2, pp. 301–312; 1888.

[6] P. Lenard, "Erzeugung von Kathodenstrahlen durch ultraviolettes Licht," *Ann. der Phys.*, vol. 2, no. 6, pp. 359–375; 1900.

[7] J. J. Thomson, "On the masses of the ions in gases at low pressures," *Phil. Mag.*, vol. 48, pp. 547–567; December, 1899.

[8] J. Elster and H. Geitel, "Notiz ueber die Zerstreuung der negativen Electricitat durch das Sonnen und Tageslicht," *Ann. der Phys.*, vol. 38, no. 9, pp. 40–41; 1889.

[9] J. Elster and H. Geitel, "Ueber die Verwendung des Natriumamalgames zu lichtelectrischen Versuchen," *Ann. der Phys.*, vol. 41, no. 10, pp. 161–176; 1890.

[10] J. Elster and H. Geitel, "Notiz ueber eine neue Form der Apparate zur Demonstration der lichtelectrischen Entladung durch Tageslicht," *Ann. der Phys.*, vol. 42, no. 4, pp. 564–567; 1891.

[11] J. Elster and H. Geitel, "Beobachtungen des atmospharischen Potentialgefalles und der ultravioletten Sonnenstrahlung, *Ann. der Phys.*, vol. 48, no. 2, pp. 338–373; 1893.

[12] J. Elster and H. Geitel, "Ueber die Vergleichung von Lichtstarken auf photoelectrischem Wege," *Ann. der Phys.*, vol. 48, no. 4, pp. 625–635; 1893.

[13] A. Einstein, "Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt," *Ann. der Phys.*, vol. 17, no. 6, pp. 132–148; 1905.

surface and is called the work function of the surface ϕ_p. Hence, $E = h\nu - \phi_p e$ where e is the charge carried by one electron.

The immediate implication of this equation is that no current will be produced for light of frequency such that $h\nu$ is less than $\phi_p e$. This conclusion has been verified in many experiments, although in recent years some modification of the interpretation of ϕ_p has been required. For pure metals, ϕ_p has been found to be related to the position of the metal in the periodic table being, in general, small for the alkali metals, and decreasing with increasing atomic weight. ϕ_p has also been shown to be equal or very nearly equal to ϕ_r, the thermionic work function involved in Richardson's equation for the emission of electrons from heated bodies. The surface photoelectric effect has proved to be basic in the development of the quantum theory of light as well as in the development of modern theories of the structure of metals. The large number of photoelectrons released from metals is attributed to the presence of "free" electrons within the metal which are relatively free to move within the metal lattice. These electrons are responsible for the high conductivity of metals. A smaller surface photoelectric effect is also found for insulators but the absence of conductivity makes it difficult to obtain a continuous emission since the insulator eventually assumes a positive charge. The reaction of the light quantum with the electron is now thought to be the primary mechanisn involved in the "photoconductive" as well as in the "photovoltaic" effect.

All three of these effects are now incorporated in commercial photoelectric cells. Their respective advantages and limitations have resulted in widely different fields of application. The current sensitivity of the photoconductive cell considerably exceeds that of the other types but disadvantages such as a large time lag have limited its use. Since no external source of voltage is required with the photovoltaic cell, one variety of this type, "the barrier-layer" or "blocking-layer" cell has found wide use in connection with low-resistance current meters in light-measuring devices such as photographic exposure meters. The photoemissive tube, or phototube, on the other hand is a high-resistance device ideally suited for use with a modern thermionic amplifying tube either of the high-vacuum or gas-filled type, the combination being capable of controlling large amounts of power. Close relatives of the phototube are the electronic "pickup" tubes used in television; these utilize large multi-element photoemissive surfaces.

The Photosensitivity of Pure Metals

Returning to the investigations of Hallwachs,[5] it will be remembered that the initial experiments on the surface photoelectric effect were made in air and that in this case ultraviolet light was required to produce the electric current. As further experiments were made and as an explanation of the effect was developed based on Einstein's equation, the wavelength threshold, that is, the longest wavelength radiation which will liberate photoelectrons, was moved into the visible region of the spectrum. As the metals which possess the lowest surface work function are sensitive to visible light, continual attempts were made to obtain surfaces of lower work function. Since a low work function has been found to be intimately associated with high chemical activity, the alkali metals which possess these properties must be examined under the best vacuum conditions. The difficulties which beset the early experimenters must be considered as being tremendous compared with experiments made today with modern vacuum pumps. However, the alkali metals were soon found to be the materials most photoelectrically sensitive to visible radiation.

The early investigators believed that the presence of gas might even be required for the passage of the photocurrent. Elster and Geitel[11,12] indeed found that a maximum current could be obtained if the envelope contained hydrogen at a certain optimum pressure. Such a device may be said to be the first gas-filled phototube. The photoelectric effect, however, was eventually found to be independent of anything but the nature of the metal surface and the character of the incident radiation. However, a subsequent amplification of this current could be obtained by ionization of a gaseous atmosphere included in the envelope. Since the active sensitive surface would be destroyed by anything other than an inert gas, such gases are always employed in gas-filled phototubes. Argon is most suitable since it possesses a low ionization potential and is relatively cheap. The value of ϕ_p for a given metal differs widely, however, depending upon the condition of the surface. Contaminations such as occluded gases or the presence of thin oxide layers change the work function markedly. The values of ϕ_p for many gas-free pure-metal surfaces have gradually been established and the maximum wavelengths for their photoelectric emission measured. The response of a metal surface to light of frequency greater than the threshold value increases until a maximum is reached. This effect was formerly attributed to the loss of energy by electrons coming from deep within the metal but is now believed to be due to the distribution of energy among the free electrons of the metal, the faster electrons being more readily emitted. Wentzel,[14] followed by others,[15–19] has computed the form of the spectral-distribution curve to be expected based on wave mechanics and a Fermi-

[14] G. Wentzel, in "Probleme der Modernen Physik," Sommerfeld Festschrift, p. 79; 1928.

[15] W. V. Houston, *Rev. Mod. Phys.*, vol. 5, p. 41; January, 1933.

[16] I. Tamm and S. Schubin, "Zur theorie des Photoeffektes an Metallen," *Zeit. für Phys.*, vol. 68, pp. 97–113; March 11, 1931.

[17] H. Frohlich, "Zum Photoeffekt an Metallen," *Ann. der Phys.*, vol. 7, pp. 103–128; October 23, 1930.

[18] W. G. Penney, "The photoelectric effect in thin metallic films," *Proc. Roy. Soc.*, vol. A133, pp. 407–417; October, 1931.

[19] K. Mitchell, "The theory of the surface photoelectric effects in metals," *Proc. Roy. Soc.*, vol. A146, pp. 442–464; September, 1934.

Dirac distribution of the energies of the free electrons. Fair agreement with experiment has been obtained. The spectral-distribution curves for the alkali metals[20] are shown in Fig. 1, which gives the relative spectral response as a function of the wavelength of the incident light. Caesium may be seen to possess the longest wavelength sensitivity being sensitive even to red radiation. For the measurement of ultraviolet radiation, various metal surfaces whose thresholds lie within that region have been used. It is frequently advisable that an ultraviolet sensitive phototube be insensitive to visible radiation; hence, the use of such metals as tantalum,

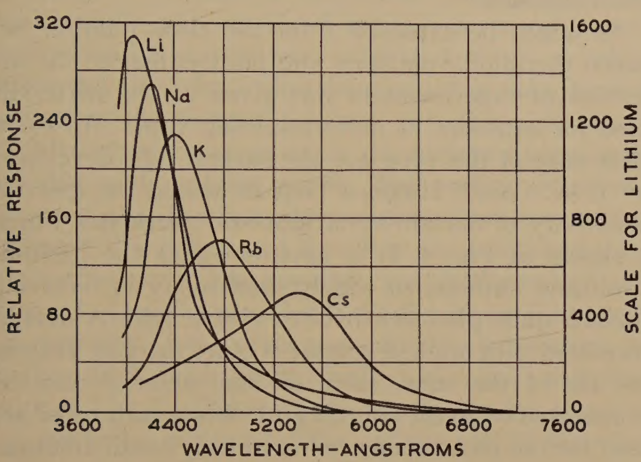


Fig. 1—Spectral-distribution curves of the alkali metals (Miss Seiler).

titanium, and thorium,[21] the spectral curves for which are shown in Fig. 2.

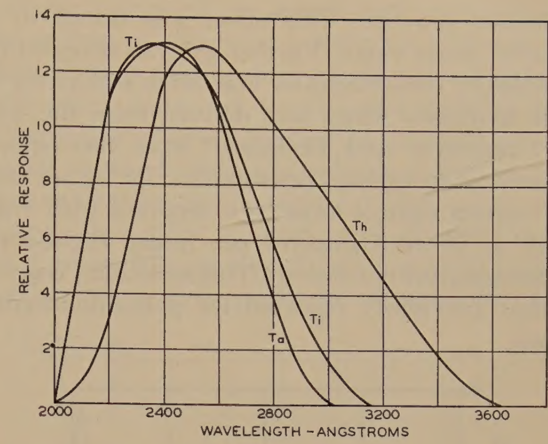


Fig. 2—Spectral distribution of tantalum, titanium, and thorium in ultraviolet-transmitting glass (Courtesy of Lamp Division, Westinghouse Electric and Manufacturing Co.).

practically all commercial phototubes were of this general type. In Fig. 3 are shown the spectral-sensitivity curves for potassium (K) and potassium hydride (K-H) as the surface is commonly called. The maximum sensitivity of the potassium-hydride surface is located in the blue portion of the spectrum, this maximum occurring at 4350 angstroms, the long-wavelength threshold occurring at 5900 angstroms. A sensitivity of 1 to 2 microamperes per lumen to a tungsten incandescent light operating at a color temperature of 2870 degrees Kelvin is obtained. Gas filling is employed to increase this value about ten to twenty times.

Very recently Tykociner, Kunz, and Garner[29] have reported the development of an improved form of the

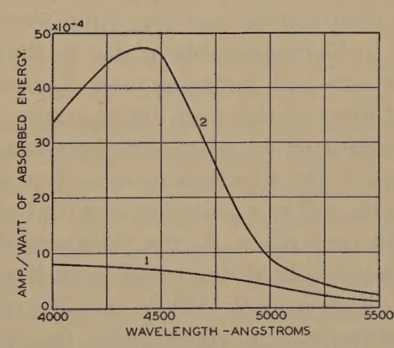


Fig. 3—Sensitization of potassium by glow discharge in hydrogen. Curve 1—potassium before sensitization. Curve 2—potassium sensitized with hydrogen (Campbell and Ritchie).

Development of Sensitized Surfaces

The first successful attempt to produce an artificially enhanced photoelectric emission was that of Elster and Geitel[22–24] who found that upon filling a potassium cell with hydrogen, the photoemission was completely destroyed but that after the passage of a glow discharge in the hydrogen atmosphere, a considerable increase in sensitivity over that of pure potassium was obtained. The hydrogen was then pumped out and argon introduced. Various improvements on this process have been made[25–28] and indeed until about 1925,

potassium-hydride surface. A reaction with carefully purified hydrogen was found to be incapable of increasing the sensitivity of a potassium tube even when the tube was submitted to a glow discharge according

[20] E. F. Seiler, "Colour sensitiveness of photoelectric cells," *Astrophys. Jour.*, vol. 52, pp. 129–153; October, 1920.

[21] H. C. Rentschler, D. E. Henry, and K. O. Smith, "Photoelectric emission from different metals," *Rev. Sci. Instr.*, vol. 3, pp. 794–802; December, 1932.

[22] J. Elster and H. Geitel, "Ueber gefarbte Hydride der Alkalimetalle und ihre photoelektrische Empfindlichkeit," *Phys. Zeit.*, vol. 11, pp. 257–262; April, 1910.

[23] J. Elster and H. Geitel, "Bemerkungen ueber die Natur der durch elektrische Entladungen auf Alkalimetallen in Vacuumrohren gebildeten farbigen Schichten," *Phys. Zeit.*, vol. 11, pp. 1082–1083; November, 1910.

[24] J. Elster and H. Geitel, "Weitere Untersuchungen an photoelektrischen Zellen mit gefarbten Kaliumkathoden," *Phys. Zeit.*, vol. 12, pp. 609–614; August, 1911.

[25] R. Pohl and P. Pringsheim, "Bemerkung ueber die lichtelektrischen effekte an kolloidalen Alkalimetallen," *Verh. der Deutsch Phys. Ges.*, vol. 13, pp. 219–223; March, 1911.

[26] R. Pohl and P. Pringsheim, "Der selektive Photoeffekt bezogen auf absorbierte Lichtenergie," *Verh. der Deutsch Phys. Ges.*, vol. 15, pp. 173–185; March, 1913.

[27] W. B. Nottingham, "Manufacture of potassium-hydride photoelectric cells," *Jour. Frank. Inst.*, vol. 205, pp. 637–648; May, 1928.

[28] N. R. Campbell, "Wasserstoff und die photoelektrische Emission aus Kalium," *Phys. Zeit.*, vol. 30, pp. 537–538; September, 1929.

[29] J. T. Tykociner, J. Kunz and L. Garner, "Photoelectric sensitization of alkali surfaces by means of electric discharges in water vapor," University of Illinois, Bulletin Series No. 325; 1940.

to common practice. This effect was traced to the absence of water vapor. Further analysis revealed that the necessary constituent for maximum sensitivity was atomic hydrogen which was derived from the water vapor. Suhrman and Theissing[30] have also reported the same observation. Sensitivities for tubes made under careful control have been reported with values as high as 10 microamperes per lumen (light-source color temperature not stated). This sensitivity is several fold that previously reported for potassium-hydride surfaces.

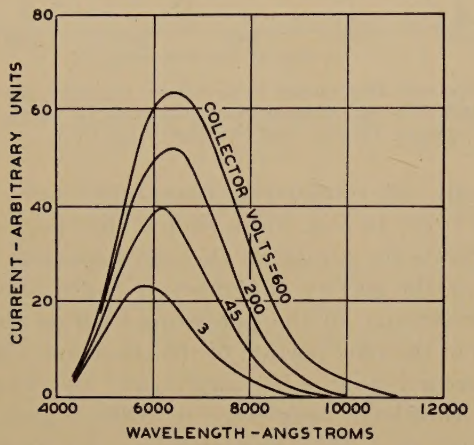


Fig. 4—Spectral-distribution curve of oxide-coated cathode as a function of the collector voltage (W. S. Huxford).

During the preparation of alkali-metal tubes, the presence of the volatile alkali vapor results in a condensation of a thin film of the metal on all the exposed parts of the tube such as the glass press or even the anode. The photocurrents arising from these films are often found to be comparable in size to the main photocurrent. Ives[31] and Suhrman and Theissing[32] have investigated films of this sort, for example, potassium or sodium deposited on platinum. Ives found that as the thickness of the film was increased, a shift in the threshold to the red to a point beyond that of the alkali metal in bulk took place. As the thickness was further increased, the threshold decreased until it approached that of the alkali in bulk. While surfaces of this type have had no commercial significance, the study of the effect has led to a better understanding of the modern thin-film surface to be discussed shortly. Ives and Olpin[33] determined the maximum excursion of the long-wavelength threshold to be equal to the first line of the primary series or resonance line of the element in question. Although this correlation has been shown to be true for the alkali metals, it has not contributed materially to an understanding of the effect. Later measurements by Brady[34] indicate that such a correlation may be invalid. The spectral selectivity found with surfaces of this sort has recently been interpreted by Ives and Briggs[35] in terms of the optical constants of the alkali-metal films. The maximum sensitivity is found to occur at that wavelength at which the electromagnetic-energy density at the surface in the standing wave which is set up in the film by interference is a maximum. The necessity for the dual relation between the wave theory and the corpuscular theory of light is nowhere more clearly required than in this aspect of photoemission.

As might be expected from the close relation between thermionic emission and photoemission, the attention of experimenters was given to the materials used for cathodes in radio receiving tubes. An early phototube of this type was the barium tube developed by T. W. Case.[36] Huxford[37] has measured the spectral sensitivity of barium-strontium-oxide cathodes which is shown in Fig. 4. It is apparent that the barium-strontium cathode so widely used today in receiving tubes is quite photosensitive to visible light. A shift in threshold with applied voltage is quite marked. During the life of the tube, some of this material may be evaporated on to the control grid. When such tubes are used for the measurement of extremely small amounts of current, the light from the filament may create a troublesome source of undesirable grid current.

The investigations of Langmuir and Kingdon[38] showed that a thin film of a volatile metal, such as caesium, may be bound to a base metal such as tungsten at temperatures far above that at which the caesium would normally be evaporated from caesium metal in bulk. Such films were found to possess a very high thermionic emission. It was further found that oxidation of the tungsten before deposition of the caesium intensified the binding forces. The value of the thermionic work function for caesium on oxygen on tungsten is given by Dushman[39] as being $\phi_T = 0.695$ volt. While this value is not necessarily equal to the photoelectric threshold, the low value of ϕ_T would indicate the probability of high photoelectric sensitivity. A photosurface of similar nature was developed by Zworykin and Wilson[40] by the deposition of

[30] R. Suhrmann and H. Theissing, "Influence of hydrogen on the photoelectric emission of potassium," *Zeit. für Phys.*, vol. 52, no. 7–8, pp. 453–463; 1928.

[31] H. E. Ives, "Photoelectric properties of thin films of alkali metals," *Astrophys. Jour.*, vol. 60, pp. 209–230; November, 1924.

[32] R. Suhrmann and H. Theissing, "Selective photoelectric effect with thin layers of potassium absorbed on polished platinum," *Zeit. für Phys.*, vol. 55, no. 11–12, pp. 701–716; 1929.

[33] H. E. Ives and A. R. Olpin, "Photoelectric long wave limit of the alkali metals," *Phys. Rev.*, [2], vol. 34, pp. 117–128; July, 1929.

[34] J. J. Brady, "The photoelectric properties of alkali metal films as a function of their thickness," *Phys. Rev.*, [2], vol. 41, p. 613; September, 1932.

[35] H. E. Ives and H. B. Briggs, "Optical properties and photoelectric emission in thin alkali metal films," *Jour. Opt. Soc. Amer.*, vol. 28, pp. 330–338; September, 1938.

[36] T. W. Case, "New strontium and barium photoelectric cells," *Phys. Rev.*, [2], vol. 17, pp. 398–399; March, 1921.

[37] W. S. Huxford, "Effect of electric field on emission of photoelectrons from oxide cathodes," *Phys. Rev.*, [2], vol. 38, pp. 379–395; August, 1931.

[38] I. Langmuir and K. H. Kingdon, "Thermionic effects caused by vapors of alkali metals," *Proc. Roy. Soc.*, vol. A107, pp. 61–79; January, 1925.

[39] K. Dushman, "Thermionic emission," *Rev. Mod. Phys.*, vol. 2, pp. 381–476; January, 1930.

[40] V. K. Zworykin and E. D. Wilson, "Caesium-magnesium photocell," *Jour. Opt. Soc. Amer.*, vol. 19, pp. 81–89; August, 1929.

caesium upon a magnesium base. The excess caesium was then driven off by heating and the resultant photosurface was found to have a high sensitivity with a peak at 4850 angstroms. The maximum response was about 2 microamperes per lumen to a tungsten lamp at 2870 degrees Kelvin. The spectral response is shown in Fig. 5. Gas filling has usually been employed to increase the sensitivity. It may be noted that this surface could withstand a temperature of 150 degrees centigrade and that its life is indefinite under very severe operating conditions.

At about the same period, Bainbridge and Charlton[41] developed a surface of a somewhat similar nature in which silver was employed as the base metal. It was later found if oxygen was admitted to the envelope after the silver layer was prepared, the tube heated, and then the oxygen pumped out, that subsequent distillation of caesium on to the surface gave a more reliable and uniform result. The surface is usually denoted as a caesium-silver surface since no reaction between the oxygen and the silver was promoted other than absorption. The spectral curve is shown in Fig. 6, curve 1. Development by Koller[42] and by Campbell[43] showed that if the silver was superficially oxidized by a glow discharge in an atmosphere of oxygen at a pressure of 1 or 2 millimeters the sensitivity was found to be markedly increased. This surface will be identified

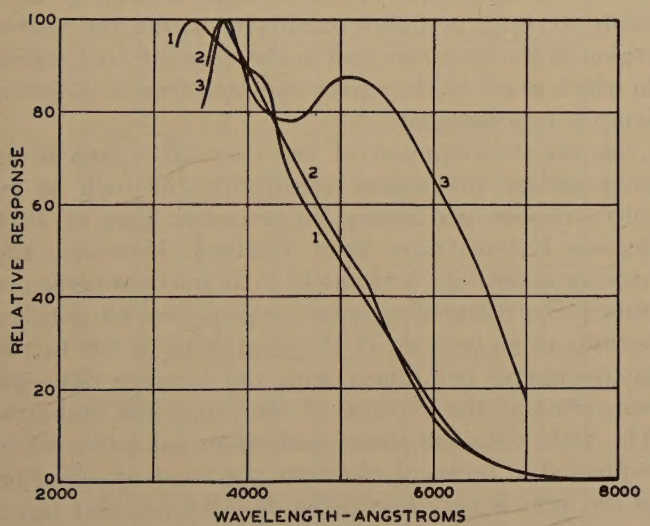


Fig. 5—Spectral-distribution curves for several sensitized surfaces. Curve 1—caesium-magnesium (Zworykin). Curve 2—sodium-sulphur (Olpin). Curve 3—sodium-oxygen-sulphur (Olpin). (Curve 1 used by permission of John Wiley and Sons, Inc.)

subsequently as caesium-oxygen-silver, including the many modifications which may be obtained by varying the proportions of the constituents. A peak in the spectral sensitivity was then found in the red and near infrared. The long-wavelength limit was extended to 10,000 to 12,000 angstroms, and sensitivities as high as 30 microamperes per lumen have been obtained. This surface is the most widely used commercially today. Various modifications of the process have been made which have led to an understanding of the mechanism involved. A solid-silver cathode or a copper cathode on which silver has been electroplated usually serves as the base for the surface. Other metals may be employed but silver is found to be the most satisfactory. Campbell found that the surface could be further sensitized by a glow discharge in argon for a short

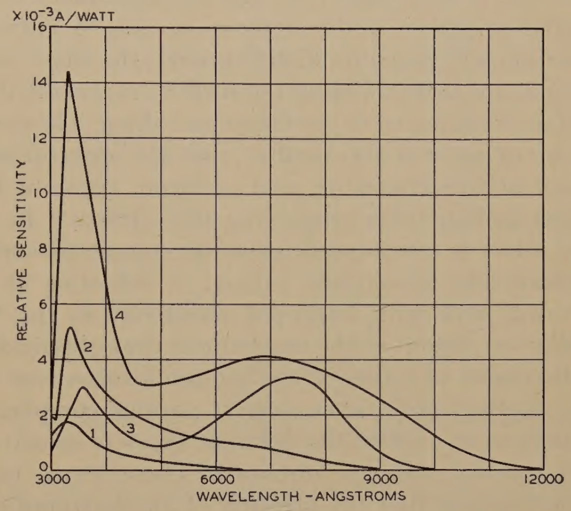


Fig. 6—Spectral-distribution curves of sensitized caesium surfaces. Curve 1—caesium-silver (Koller). Curve 2—caesium-oxygen-silver (deficiency of caesium) (Campbell and Ritchie). Curve 3—caesium-oxygen-silver (excess of caesium) (Ritchie). Curve 4—silver-caesium-oxygen-silver (Asao and Suzuki).

period of time, although continued bombardment would destroy the surface. Prescott and Kelly[44] have published a detailed account of one procedure giving very satisfactory results in which great care was taken to obtain a careful balance between the amount of oxygen and the amount of caesium employed.

N. R. Campbell[43] established that the amount of caesium required to obtain maximum sensitivity was proportional to the amount of oxidation of the silver. Two types of such surfaces are shown in Fig. 6, curves 2 and 3, in which the ratio of caesium to oxygen is larger for the surface of curve 3. Campbell believes that surfaces of this type may be considered as consisting of a layerlike structure in which a thin film of alkali, probably monatomic in nature, is superimposed on a matrix of an oxide of caesium which is in turn built upon the silver base by reduction of the silver oxide. The nature of the oxide of caesium is rather indeterminate. Probably the intermediate layer is a mixture of various oxides of caesium, unreduced silver oxide, and free silver. According to deBoer and Teves[45] the presence of a certain amount of free silver enhances the

[41] Bainbridge and Charlton, see reference 42, page 1639.

[42] L. R. Koller, "Photoelectric emission from thin films of caesium," *Phys. Rev.*, [2], vol. 36, pp. 1639–1647; December, 1930.

[43] N. R. Campbell, "Photoelectric emission of thin films," *Phil. Mag.*, vol. 12, pp. 173–185; July, 1931.

[44] C. H. Prescott, Jr. and M. J. Kelly, "The caesium-oxygen-silver-photoelectric cell," *Bell Sys. Tech. Jour.*, vol. 11, pp. 334–367; July, 1932.

[45] J. H. deBoer and M. Teves, "Electron Emission and Absorption Phenomena," The Macmillan Company, Cambridge, England, 1935, p. 330.

sensitivity by providing a source of conductivity in an otherwise poorly conducting layer. They assume that the electrons emitted are the result of a photoionization of the surface caesium atoms. By a partial reduction of the silver oxide in a hydrogen atmosphere, they obtain the presence of a certain amount of free silver. By such a technique, surfaces have been obtained in which the sensitivity extends into the infrared region to 17,000 angstroms. Asao and Suzuki[46] have shown that it is possible to enhance the sensitivity by evaporating a thin film of silver upon the surface subsequent to the normal procedure outlined above (see Fig. 6, curve 4). Caesium will probably distill through the silver to re-form a monatomic film on the surface, especially if the surface is subjected to a subsequent baking. The evaporation of silver is also used to alter the spectral sensitivity of the iconoscope and orthicon, types of television pickup tubes employing this surface.[47] In this case silver is evaporated upon the completed surface without any subsequent baking. A reduction of the infrared peak with improved sensitivity in the blue and green regions of the spectrum is thus obtained. In a discussion of forms of the caesium-oxide-silver surface, deBoer and Teves suggest various complicated structures to explain the different types of sensitivity and spectral response obtained. There seems to be little question that the presence of an electronegative element is the essential constituent in producing a reduction of the work function of the metal surface. Others have substituted sulphur for the oxygen with somewhat similar results, although the sensitivity reported has not been as high. Campbell interprets the sensitivity in terms of a transmission of electrons through a surface potential barrier consisting of a series of maxima and minima in the potential as a function of the distance from the surface. On a wave mechanical theory, such a potential barrier will have a selective transmission for electrons of certain energies. Olpin[48] has computed the atomic spacing which would be expected to give a selective transmission on the basis of a crystalline structure, through which the electrons would pass according to the wave mechanical theory. He shows that such distances as computed from the observed wavelengths of the spectral maximum show a direct ratio to the atomic diameters of the atoms involved and that the correlation is remarkable considering the uncertainties in the theory and in the experimental quantities involved. Olpin's results are probably somewhat fortuitous as pointed out by Zachariasen.[49]

An extensive series of investigations on the sensitization of alkali surfaces has been carried out by Olpin.[50] Among various inorganic substances, Olpin tried oxygen, water vapor, and sulphur. Among the organic materials tested were benzene, acetone, and organic dyes such as are used for sensitization in photography, for example eosine, kryptocyanine, and dicyanine. Olpin found that practically every substance tried increased the sensitivity of the alkali metal to red light. He concluded that a contamination common to all the substances might be the cause and suggested water vapor. A more reasonable explanation now appears to be that the presence of any electronegative element in the surface layer is sufficient to distort the potential barrier so as to increase the threshold wavelength. Hence, oxygen or sulphur may be the most likely sources of the increased sensitivity in all the above cases. Olpin notes the effect of depositing a thin film of sodium on a surface of sodium already exposed to sulphur. A surface very similar to the caesium-oxygen-silver in structure is probably obtained. Curves showing the spectral sensitivities for some of the surfaces produced by Olpin are shown in Fig. 5. The commercial importance of these discoveries has been limited by the widespread use of the caesium-oxygen-silver type of surface, but as in the case of Ives' work on thin films, the general conclusions obtained have been very valuable in forming a picture of the type of surface most likely to result in a high sensitivity within the visible region of the spectrum and in the near-infrared region in which much of the energy radiated from a tungsten lamp is concentrated.

As the development of the caesium-oxygen-silver photosurface progressed, sensitivities as high as 50 microamperes per lumen (to tungsten light at 2870 degrees Kelvin) have been obtained. However, the average sensitivity is about 15 to 25 microamperes per lumen. Gas filling of caesium-oxygen-silver phototubes permits of an increase in the sensitivity to 100 to 150 microamperes per lumen with the increase obtained somewhat at the expense of constancy and stability. The yield obtained from caesium-oxygen-silver when expressed in terms of electrons per incident quantum of the light is very low, being only 0.6 per cent (for a sensitivity of 40 microamperes per lumen) at the peak in sensitivity in the infrared. At the peak in the near ultraviolet, the value is somewhat higher, being about 1.6 per cent.

Glover and Janes[51] have reported the development of a composite caesium photosurface in which the sensitive surface is formed upon a metal base such as nickel. The spectral maximum for this surface is found to occur at 3750 angstroms and is accompanied by a long-wavelength threshold in the neighborhood of 6300

[46] S. Asao and M. Suzuki, "Improvement of thin film caesium photoelectric tube," *Proc. Phys. Math. Soc.* (Japan), series 3, vol. 12, pp. 247–250; October, 1930.

[47] R. B. Janes and W. H. Hickok, "Recent improvements in the design and characteristics of the iconoscope," Proc. I.R.E., vol. 27, pp. 535–540; September, 1939.

[48] A. R. Olpin, "An interpretation of the selective photoelectric effect from two component cathodes," *Phys. Rev.*, [2], vol. 38, pp. 1745–1757; November, 1931.

[49] W. H. Zachariasen, "On the interpretation of the selective photoelectric effect from two component cathodes," *Phys. Rev.*, [2], vol. 38, p. 2290; December, 1931.

[50] A. R. Olpin, "Method of enhancing the sensitiveness of alkali metal photoelectric cells," *Phys. Rev.*, [2], vol. 36, pp. 251–295; July, 1930.

[51] A. M. Glover and R. B. Janes, "A new high-sensitivity photosurface," *Electronics*, vol. 13, pp. 26–27; August, 1940.

angstroms. Sensitivity as high as 90 microamperes per lumen (to tungsten light at 2870 degrees Kelvin) has been obtained which corresponds to a quantum efficiency as high as 26 per cent at the wavelength of peak sensitivity. This surface is found to be very stable at conditions under which the caesium-oxygen-silver surface is subject to loss of sensitivity. The spectral sensitivity for this surface enclosed in a lime-glass bulb is shown in Fig. 7. The sensitivity in the ultraviolet region between 2000 and 3000 angstroms is undoubtedly higher than for any surface yet available.

Görlich[52] has reported the development of a transparent surface in which the sensitivity is largely limited to the blue and green region of the spectrum, but for which the yield is remarkably high in this region. The surface is obtained by the deposition of a layer of antimony upon the glass envelope. The antimony is then treated with oxygen and subjected to caesium vapor. This surface may be deposited in a transparent layer sufficiently thin so that the sensitivity may be obtained by light incident upon the surface from the side opposite to that from which the current is collected. Görlich reports that the selective maximum lies between 4000 and 6000 angstroms and that the position of the maximum is largely unaffected by the substitution of sodium, potassium, rubidium, or lithium in place of the caesium. Gopstein and Khorosh[53] report measurements on such surfaces and upon surfaces in which the antimony is replaced by bismuth and arsenic for illumination both from the front and from the rear of the surface. They quote values for the sensitivity as:

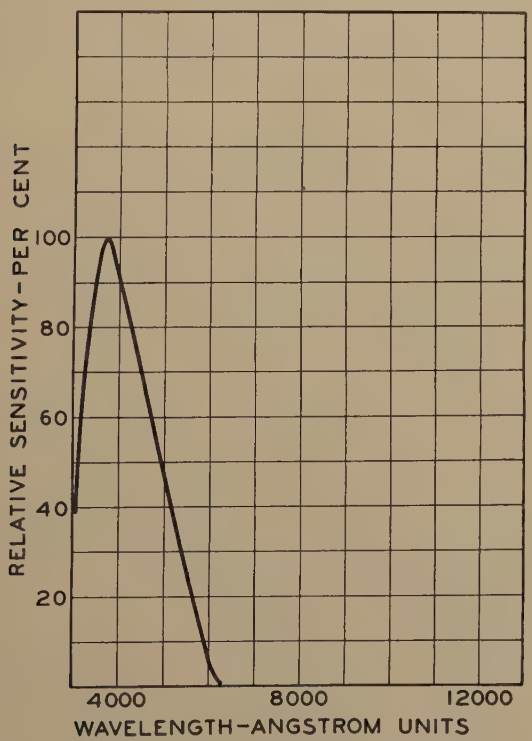


Fig. 7—Spectral-distribution curve of composite caesium surface in lime-glass bulb (Glover and Janes).

antimony-caesium	60–70 microamperes per lumen
bismuth-caesium	10–15 microamperes per lumen
silver-caesium	20–40 microamperes per lumen

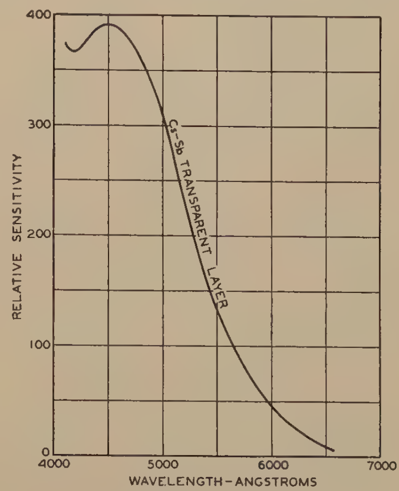


Fig. 8—Spectral-distribution curve of transparent caesium-antimony alloy (Görlich).

The color of these surfaces is a burnt red, indicating a possible correlation between the photoemission and the optical absorption of the surface.[54] Further development[52] has included the sensitization of these surfaces by controlled amounts of oxygen to produce a shift of the wavelength of the peak sensitivity towards the red end of the spectrum. The spectral-sensitivity curve for the caesium-antimony surface reported by Görlich is shown in Fig. 8.

Technical Development

The development of commercial phototubes has, of course, accompanied the development of radio receiving tubes. Thus, improvements in the latter, both of technical and of economic significance, have been largely incorporated in the manufacture of phototubes. Some of these are so obvious as to need no mention but others may be of some interest. In general, it may be said that the surface of a phototube is more sensitive to contamination than the cathode of a receiving tube. The cathode surface of a receiving tube is believed to be covered with a thin film of barium which is replenished during the life of the tube by transfer of barium toward the surface of the cathode by electrolysis. The result is that as traces of residual gas react with the metallic barium on the surface, this gas is gradually cleaned up until insufficient to do any further harm. On the other hand, the alkali elements which are usually present in phototubes and which are also very active chemically react during processing

[52] P. Görlich, "Sensitization of transparent compound photocathodes," *Zeit. für Tech. Phys.*, vol. 18, pp. 460–462; November, 1937.

[53] N. M. Gopstein and D. M. Khorosh, "Photoeffect and secondary emission with alloy cathodes," *Jour. Tech. Phys.* (U.S.S.R.), vol. 8, pp. 2103–2106; 1938.

[54] P. Lukirsky and N. N. Lusheva, "Photosurfaces with high selective sensitivity," *Jour. Tech. Phys.* (U.S.S.R.), vol. 7, pp. 1900–1904; 1937.

with residual gases present in the tube or released from the walls and parts by electron bombardment. No electrolysis of alkali atoms toward the surface takes place subsequent to formation of the surface. Hence, if the surface should be contaminated there would be no additional supply of alkali available to replenish the surface atoms. The sensitivity obtained thus would be decreased. As a consequence, phototubes require special care on exhaust procedure. A getter is not generally used because of the high activity of the alkali elements since the free alkali present during the formation of the surface provides the final gettering. On the other hand, once a sensitive surface has been formed, its life is practically indefinite since the cathode is operated at low temperature with no loss of material by evaporation. If an inert gas is used to provide amplification of the electron current, it must be one of the noble gases, such as helium, argon, or neon. The gas is usually purified by an electrical discharge to misch-metal or some other active surface in a bulb containing the supply of gas. Argon is usually used because of its low cost and relatively low ionization potential. Vacuum diffusion pumps of good quality are universally used in conjunction with liquid-air traps to remove traces of water vapor and other condensable gases.

If the active surface is a pure metal, this may be prepared by cathode sputtering or by evaporation of the metal in vacuo. The latter method is being used widely for the preparation of metal surfaces sensitive to the ultraviolet. In early phototubes in which the presence of an alkali metal was desired, the alkali was usually obtained by distillation from bulk metal contained in a side tube. This method is tedious and unsuited to large-scale production. Burt[55] found that sodium could be electrolytically formed upon the inner surface of a soda-glass bulb, if the bulb were placed in an electrolytic bath which would furnish the sodium ion to the glass. Zworykin used a potassium-containing glass for the purpose of obtaining a potassium surface in similar fashion. In recent years, the alkali metal has been obtained by the reduction of a suitable salt. A reducing element is provided with the salt incorporated in a pellet which is brought to high temperature usually by high-frequency induction. For obtaining caesium, it is usual to employ a mixture of caesium dichromate and silicon. An alternative mixture is a combination of caesium chromate, aluminum, and tantalum. This method is convenient and provides a controlled amount of the desired alkali directly within the tube. Hydrogen which is required for the potassium-hydrogen cell may be obtained by heating palladium in a side tube or by dosing from a bottle of gas connected to the exhaust system.

In the preparation of the caesium-oxygen-silver or similar surfaces, an excess of alkali is useful in the preparation of the surface. This excess must then be removed in order to prevent electrical leakage across any films condensed upon the stem press or interior of the bulb surface. The removal may be accomplished by the introduction of a compound such as oxidized copper or stannic oxide or by the reaction of the excess caesium with the lead of the glass stem press. Since this reaction causes a darkening of the glass, the bulbs of alkali-containing phototubes must be of lead-free glass such as lime glass or one of the borosilicate glasses.

For convenience, it is usually desirable that the cathode surface be formed upon a metal element which may be directly connected to one of the lead-in wires in the stem press. This metal support may be of nickel, copper, or even be of solid silver. The last is used frequently in the preparation of the caesium-oxygen-silver surface. The anode is usually a nickel wire also directly connected to a lead-in wire in the press and, then, to one of the base pins. For many purposes, however, it is desirable to provide an exceptionally high-resistance path between anode and cathode. This is accomplished by sealing the anode connection through the glass envelope separately. A guard ring consisting of a strip of metallic paste painted on the glass between the anode connection and the cathode connection is also of use in by-passing undesirable leakage currents from the measuring circuit. Surface leakage may also be reduced by coating the bulb surface with a nonhygroscopic wax such as ceresin.

For many years, applications of phototubes were limited to the scientific laboratory because of the small size of the currents obtained and because of the delicate nature of the instruments required for measuring such currents. The development of the vacuum tube, however, has changed the situation completely. Because of the high-resistance characteristic of phototubes, the small current output of a phototube may be converted into a voltage of sufficient size to be used as the grid voltage of a commercial radio tube. No extraordinary precautions need be taken other than to insure that base and socket leakages be kept to a minimum and that a suitable amplifying tube be used. (A tube with a top grid cap is normally desirable because of the increased grid resistance thus provided). For the detection or measurement of extraordinarily small amounts of light, amplifying tubes have been developed in which all sources of extraneous grid current have been minimized by providing high internal-leakage paths and by adjusting the operating characteristics so that the applied voltages are insufficient to ionize any residual gas present in the tube after manufacture.[56,57] These tubes have largely replaced the delicate laboratory electrometer and are, therefore, frequently referred to as electrometer tubes. Balanced bridge circuits[58] have also been employed to decrease the effect

[55] R. C. Burt, "Sodium by electrolysis through glass," *Jour. Opt. Soc. Amer.* and *Rev. Sci. Instr.*, vol. 11, pp. 87–91; July, 1925.

[56] G. F. Metcalf and B. J. Thompson, "A low grid current vacuum tube," *Phys. Rev.*, vol. 36, pp. 1489–1494; November, 1930.

[57] H. Nelson, "A vacuum tube electrometer," *Rev. Sci. Instr.*, vol. 1, pp. 281–284; May, 1930.

[58] L. A. DuBridge and H. Brown, "An improved d.c. amplifying circuit," *Rev. Sci. Instr.*, vol. 4, pp. 532–536; October, 1933.

of changing conditions of light and fluctuations due to changes in battery or line voltages.

The limited sensitivity of the phototube may be partially overcome by gas filling. Campbell[59] has studied the performance of tubes for varying gas pressures. He shows that the voltage at which the tube breaks down, that is, the voltage at which the ionization becomes cumulative, is a function of both the amount of light and the pressure of gas used. The effect of the gas upon the frequency response of such tubes is shown in Fig. 9. A portion of this effect has been attributed to the finite time required for positive ions of the inert gas to reach the cathode. Also, the positive ions on reaching the cathode, release secondary electrons which contribute to the current. The amount of this secondary emission is quite high for the caesium-oxygen-silver surface as compared with the potassium-hydrogen surface. A more rapid increase in current with an increase in applied voltage is thereby found (see Fig. 10). Kingdon and Thompson[60] attributed the major portion of the effect to the time required for neutralization of the positive ion after it reaches the cathode. They assumed that during this interval, the work function of the cathode will be lowered and that increased secondary emission would be expected. Skellet[61] has shown a definite correlation between the time lag and the geometry of the tube, indicating the importance of the first factor. More recently, Huxford[62] and Kruithof[63] have shown that the presence of metastable atoms can explain the excessive lag at low frequencies. The increased lag when helium is used is then readily explained since the probability of excitation of metastable atoms is known to be greater for helium than for argon. While the poorer frequency response of a gas-filled tube, as compared with a high-vacuum tube, is annoying, the falling off at higher frequencies is readily compensated by adjustment of the amplifier characteristics. The effect, therefore, has been no bar to the use of such tubes in the motion-picture industry.

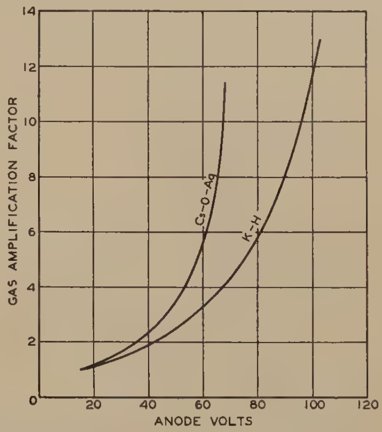


Fig. 10—Amplification by gas filling with caesium-oxygen-silver and potassium-hydride surfaces (Campbell and Ritchie).

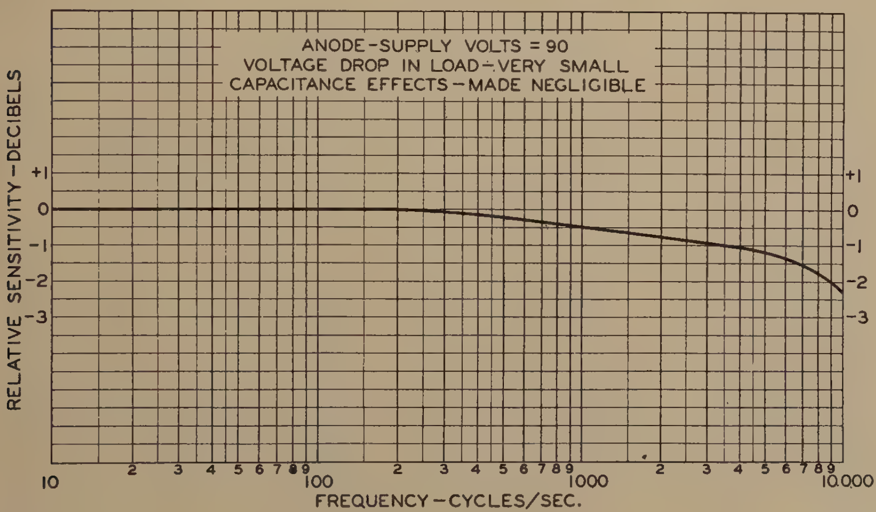


Fig. 9—Frequency response of gas-filled phototube (Courtesy of RCA Manufacturing Co.).

In an attempt to avoid the disadvantages which accompany the use of gas to increase the sensitivity, amplification by the use of secondary-emission currents released by electron bombardment of auxiliary electrodes has been employed. Secondary emission is an effect whose presence has been noted in receiving tubes for many years. In general, the secondary emission obtained from pure metals is low; however, at a few hundred volts, release of as many as 6 to 8 electrons per primary electron has been obtained from caesium-oxygen-silver surfaces, and somewhat lower ratios are readily obtained from oxides. The emission varies rapidly with voltage and passes through a maximum for a value of a few hundred volts on the bombarded electrode. The emission is stable at the current densities encountered in phototubes but in order to obtain a maximum benefit, several stages of amplification are desirable. A scheme for directing the electron streams to successive electrodes has been devised.[64] This method employs the focusing effect of the combined action of crossed electrostatic and electromagnetic fields. A somewhat simpler

[59] N. R. Campbell and D. Ritchie, "Photoelectric Cells," Sir Isaac Pitman and Sons, Ltd., London, England, 1934, p. 66.
[60] K. H. Kingdon and H. Thompson, "Some experiments with gas-filled Cs-O-Ag photoelectric cells," *Physics*, vol. 1, pp. 343–351; 1931.
[61] A. Skellet, "Time lag in gas-filled photoelectric cells," *Jour. Appl. Phys.*, vol. 9, pp. 631–634; October, 1938.
[62] W. S. Huxford, "Townsend ionization coefficients in Cs-Ag-O phototubes filled with argon," *Phys. Rev.*, [2], vol. 55, pp. 754–762; April, 1939.
[63] A. A. Kruithof, "Time lag phenomena in gas-filled photocells," *Philips Tech. Rev.*, vol. 4, pp. 48–55; February, 1939.
[64] V. K. Zworykin, G. Morton, and L. Malter, "The secondary emission multiplier," Proc. I.R.E., vol. 24, pp. 351–375; March, 1936.

arrangement dispenses with the magnetic fields by the use of a number of secondary-emissive grids arranged in a row.[65] A large portion of the current passes directly through each grid, however, thus reducing the over-all amplification. Pierce[66] and Zworykin and Rajchman[67,68] have employed mechanical models to develop all-electrostatic focusing structures which conserve all the electrons available and give maximum amplifications. In addition, the structure of

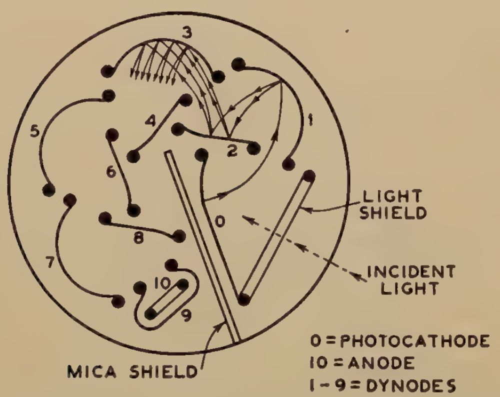


Fig. 11—Electrode structure of electrostatically focused electron multiplier (Janes and Glover).

Zworykin and Rajchman is so designed that it is difficult for positive ions, formed in the inevitable residual gas, to reach the photocathode and thus produce unwanted currents which might lead to uncontrolled regeneration. An electron multiplier employing this principle has recently been described by Janes and Glover,[69] the construction of which is shown in Fig. 11. A gain of 250,000 is readily obtainable at a total of 1250 volts between the collecting anode and the photocathode. Considerably higher figures have been obtained in laboratory models. The size of the tube is no greater than that of a simple phototube.

The phototube is the heart of most sound-picture systems. Its use is also general in "facsimile" systems which accomplish the transmission of pictures by wire or wireless. For this service, the picture is broken up into a series of elements in a fashion somewhat similar to that of a half tone and the light reflected from each element is directed successively to a phototube by some mechanical means. The time required to transmit a reproduction of a page of this size may be several minutes. For television, the requirement is, of course, to transmit the reproduction almost instantaneously. For this purpose, a two-dimensional system is required; that is, all elements of a picture or a live scene must be scanned and transmitted within a time sufficiently short for the eye to recreate by persistence of vision the original picture from light signals supplied by the reproducing mechanism to the eye. A photosurface consisting of a multitude of sensitive elements is one answer to this problem.[70] The caesium-oxygen-silver surface used is obtained by methods analogous to those previously described. However, the surface is broken up into a large number of individual elements by depositing the silver in a thin film upon a sheet of mica and heating the film until it coagulates into hemispherical lumps of silver. The process from this point on is very similar to that used in phototubes. To obtain good resolution, no excess caesium may be left upon the mica, however, so that an amount less than that required for obtaining a maximum photosensitivity is usually employed. The surface may be processed further to match its characteristics more nearly to the spectral response of the eye by the subsequent deposition of a very thin film of silver.[47] In other television pickup devices, it has been found desirable to make use of semitransparent photocathodes deposited either upon the wall of the glass envelope or upon a thin transparent sheet of mica.[71,72] The caesium-oxygen-silver surface is again most universally used, although in this case the film is usually continuous.

Little mention need be made of the countless applications of phototubes in which the various surfaces described above have been incorporated. In general, it may be said that a trend toward a more compact design is apparent. Uniformity of construction and of characteristics is a much-sought-after goal, toward which considerable progress has been made.

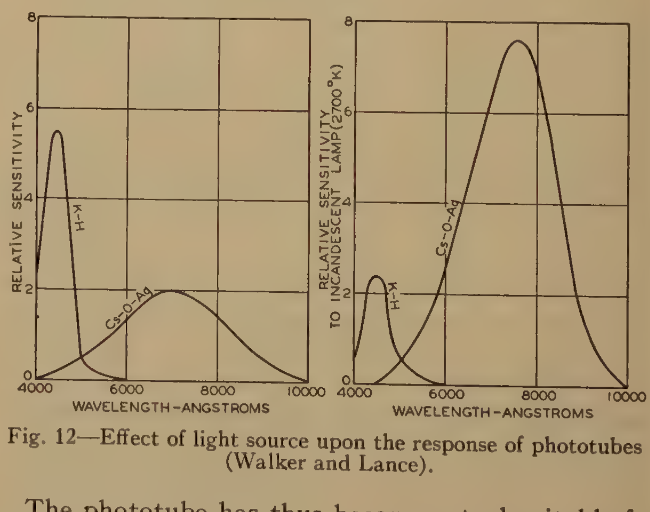


Fig. 12—Effect of light source upon the response of phototubes (Walker and Lance).

The phototube has thus become a tool suitable for commercial use as a rugged, reliable, light-sensitive

[65] W. Kluge, O. Beyer, and H. Steyskal, "Photocells with secondary emission amplification," *Zeit. für Tech. Phys.*, vol. 18, no. 8, pp. 219–228; 1937.

[66] J. R. Pierce, "Electron multiplier design," *Bell Lab. Rec.*, vol. 16, pp. 305–309; May, 1938.

[67] V. K. Zworykin and J. Rajchman, "The electrostatic electron multiplier," Proc. I.R.E., vol. 27, pp. 558–565; September, 1939.

[68] J. A. Rajchman and R. L. Snyder, "An electrically focused multiplier phototube," *Electronics*, vol. 13, pp. 20–23, 58–60; December, 1940.

[69] R. B. Janes and A. M. Glover, "Recent developments in phototubes," *RCA Rev.*, vol. 6, pp. 43–54; July, 1941.

[70] V. K. Zworykin, G. A. Morton, and L. E. Flory, "Theory and performance of the iconoscope," Proc. I.R.E., vol. 25, pp. 1071–1093; August, 1937.

[71] C. C. Larson and B. C. Gardner, "Image dissector," *Electronics*, vol. 12, pp. 24–27; October, 1939.

[72] Albert Rose and Harley Iams, "Television pickup tubes using low-velocity electron-beam scanning," Proc. I.R.E., vol. 27, pp. 547–555; September, 1939.

device. The considerable quantity of phototubes used by the motion-picture industry is evidence of this. For this application and, of course, for many others, the incandescent lamp has been the most suitable light source available. For use with this light source, the caesium-oxygen-silver surface is highly desirable, particularly if the lamp is operated at reduced voltage in order to provide increased life. On the other hand, for many applications, daylight is the source of light to be detected, and in many chemical processes or in color sorting, blue or green light is the predominant color. In late years, several new sources of illumination have appeared such as the sodium-vapor lamp,[73] the high-pressure mercury arc, and fluorescent lamps. Since the available output from the phototube at a given wavelength is the product of the emission of the source and the sensitivity of the phototube at that wavelength, the over-all sensitivity curve of the combination may be measured or computed. The effect of the light source on the response of caesium-oxygen-silver and a potassium-hydrogen cell is given in Fig. 12, showing the relatively high sensitivity of the caesium-oxygen-silver surface with an incandescent-light source. For many applications, a phototube whose sensitivity matches that of the eye would be desirable. A close approach to this requirement has been obtained with various types of barrier-layer photovoltaic cells. However, when amplification of the output current is required, this type of cell is seldom suitable. In this case, the use of a phototube, together with the proper choice of filter, will be the more satisfactory solution.

[73] L. J. Buttolph, "High efficiency mercury and sodium vapor amps," *Jour. Opt. Soc. Amer.*, vol. 29, pp. 124–130; March, 1939.

Bibliography

A complete bibliography of the development of photoemissive surfaces is, of course, out of the question. However, an attempt has been made to list sufficient references to indicate both the trend of the development, and the names of those most active in the field. From these as a start, it should be possible to trace through the literature subsequent references to any item of particular interest. There are also listed several representative texts to which reference may be made as well as comprehensive summaries which are available for a survey of the theory of the photoelectric effect. In addition, several surveys are listed on the photoconductive and photovoltaic effects which should be of value to anyone interested in the subject of photoelectricity.

(1) A. L. Hughes and L. A. DuBridge, "Photoelectric Phenomena," McGraw-Hill Book Co., Inc., New York, N. Y., 1932.
(2) N. R. Campbell and D. Ritchie, "Photoelectric Cells," Sir Isaac Pitman and Sons, Ltd., London, England, 1934.
(3) B. Lange, "Die Photoelemente und ihre Anwendung," J. A. Barth-Verlag, Leipzig, Germany, 1936.
(4) F. Seitz, "Modern Theory of Solids," McGraw-Hill Book Co., Inc., New York, N. Y., 1940.
(5) K. T. Compton and I. Langmuir, "Electrical discharges in gases," *Rev. Mod. Phys.*, vol. 2, pp. 123–242; April, 1930.
(6) L. B. Linford, "Recent developments in the study of the external photoeffect," *Rev. Mod. Phys.*, vol. 5, pp. 34–61; January, 1933.
(7) F. C. Nix, "Photoconductivity," *Rev. Mod. Phys.*, vol. 4, pp. 723–766; October, 1932.
(8) L. O. Grondahl, "The copper-cuprous oxide rectifier and photoelectric cell," *Rev. Mod. Phys.*, vol. 5, pp. 141–168; April, 1933.
(9) G. Déjardin, "Procédés modernes de fabrication des cellules photoelectriques," *Rev. Gen. dé l'Elect.*, vol. 33, p. 3, p. 36; 1933.
(10) "Photoelectric cells and their applications," (A joint discussion with many contributed papers); *Trans. Opt. Soc.*, vol. 314, pp. 233–240; 1929–1930.

Magnetic Recording and Some of Its Applications in the Broadcast Field*

S. J. BEGUN†, NONMEMBER, I.R.E.

Summary—*Broadcast stations in this country today depend upon mechanical recording as the only means of storing a program for delayed transmission or for reference purposes.*

Magnetic recording, which has been frequently used in Europe, has not as yet been introduced into this country in the broadcast field although, because of its characteristics, it has some definite advantages. The magnetic sound carrier has the outstanding feature that it can be repeatedly used for new recordings without any deterioration. On the other hand, a magnetic impression on the sound carrier is not harmed by repeated reproduction. Magnetic recording is, therefore, particularly suitable where time delay in transmission or many repetitions of the same program material are required.

This paper will deal briefly with the three essential processes of magnetic recording: the obliterating, the recording, and, finally, the reproducing of a record. The magnetic material to be used will be discussed briefly as will some other necessary electrical and mechanical requirements for obtaining a desired frequency and dynamic range.

Reference will be made to applications of magnetic recording systems by European broadcast stations for program delay. Furthermore, a unit will be described which permits short time delays and thus can be used either for producing artificial reverberation or artificial echo.

Magnetic recording, in addition to this, promises to be applicable for "spot" announcements.

* Decimal classification: 621.385.97. Original manuscript received by the Institute, March 17, 1941. Presented, Sixteenth Annual Convention, January 9, 1941.

† Brush Development Company, Cleveland, Ohio.

RECORDING may be defined as the process of leaving certain impressions—they may be mechanical, optical, magnetic, electrostatic, or chemical—on a material called the signal carrier. These impressions have some predetermined relationship to the signal being recorded and can, after a desirable time interval, be converted by suitable means into impulses which are a facsimile of the original signal.

Using an instantaneous method of recording which does not require any processing, reproduction can take place either almost immediately after the recording process, or after the lapse of a longer period of time. For this reason, the over-all process of recording and reproducing is equivalent to a time-delay circuit covering a certain frequency range with a certain frequency characteristic, and having a predetermined phase shift for the different frequencies of the band width. Recording, therefore, can be substituted for an electrical time-delay circuit and will find its particular

application where it is no longer economical to use electrical means because of the bulk of equipment involved.

In principle, any recording method can be employed for such time-delay applications. But the chosen method should depend finally upon the specific requirements, at the same time giving consideration to the expenses.

Mechanical recording and, specifically, disk recording, has been generally adopted by broadcasting sta-

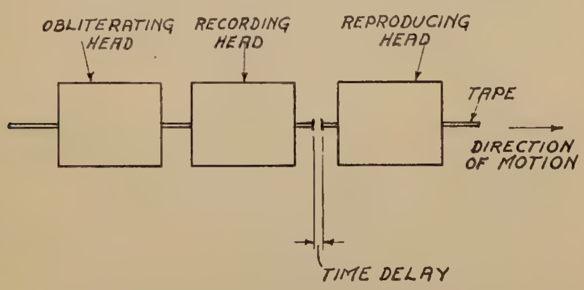


Fig. 1—Sequence of heads in magnetic recording.

tions for their particular applications. Whether certain program material has to be delayed for half an hour or whether it has to be kept on file for reference purposes, mechanical recording is used. It is quite natural to ask whether this is the most advantageous practice. This question becomes even more important in cases where the program material has to be delayed only a short period of time, and the quality of recording must be as good as possible; or, in other words, where the time-delay circuit must cover as wide a frequency range as the original program material and must have as few nonlinear elements as possible. Only a high-fidelity recording system meets these requirements. The disk recording method is, of course, capable of these high-quality requirements, but at the same time, there can be no question that if used for short time delays only, the weak point in disk recording is poor economy. The expensive cellulose-nitrate blanks can be used only once; cheaper record materials have not as yet been found to give equal performance characteristics. The question arises, therefore, whether or not there are other methods of recording which permit the frequent use of the same sound carrier, and eliminate this continuous expense. It appears that of all recording methods now in use, the only one which can meet the necessary requirements is magnetic recording.

The value of magnetic recording for time-delay applications has been recognized for a number of years by different European broadcasting companies. In this country it has thus far been used very little; as a matter of fact, it is not very familiar to the greater majority of radio and acoustical engineers. However, this recording method appears now to be gaining a foothold in this country, and it seems timely to present an outline of its operating principles.

In general, any recording system must include the processes of recording and reproduction, and, if the same signal carrier, as in magnetic recording, is to be used repeatedly, a third process of obliterating the previous signal must be incorporated. Accordingly, the signal carrier must pass in sequence three stations which are commonly known as the obliterating head, the recording head, and the reproducing head. The position of these three heads is shown in Fig. 1.

In practice, the magnetic signal carrier is a steel tape having a thickness small compared with its other dimensions. All of the heads perform a similar function in coupling the external electrical force with the internal magnetic state of the signal carrier. All have two soft-iron pole pieces disposed on opposite faces of the tape, and may or may not have coils, depending upon their function. In addition, provisions have to be made to impart motion to the signal carrier.

A good deal of literature on magnetic recording is available, but in the interests of completeness, it seems well not to omit a review of its pertinent points.

The Magnetic Process

The general principles involved in any magnetic process will now be considered. An electric current always sets up a magnetic force, measured in oersteds, which is proportional to the current. In air, this magnetic force produces by definition a magnetic field of the same number of lines as the magnetization force. In paramagnetic materials, including most of the ferric alloys and certain nickel-aluminum combinations, the induction produced has more lines than the inducing magnetic field. In every magnetic material, the relationship between the magnetic force and the resultant magnetic induction is expressed by the familiar hysteresis loop, such as shown in Fig. 2.

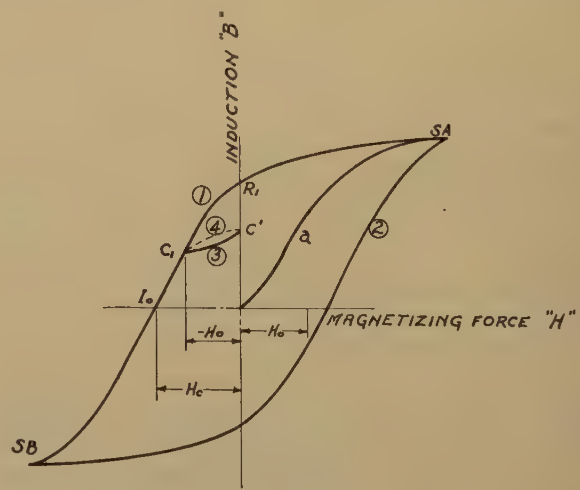


Fig. 2—Typical hysteresis loop of a paramagnetic substance.

Assuming that the magnetic material is in an unmagnetized condition, its initial induction, under an increasing magnetizing force H_o, will proceed along the branch a of the magnetizing curve. After being subjected to a magnetizing force of sufficient intensity,

the magnetic material reaches its saturation value, shown at SA, and the induction does not materially increase. Upon reversing the magnetizing force, the induction will no longer follow curve a, but instead will follow the upper branch 1 of the hysteresis loop. When the magnetizing force has reached its zero value, a magnetic induction R_1, called the residual induction, remains in the magnetic material. To bring the material to the point I_o where its induction is zero, a magnetizing force H_c, opposite to the original magnetizing force and known as the coercive force, has to be applied. On further increasing this reversed magnetic field, the material again becomes saturated at point SB with opposite magnetic polarity. By reversing the direction of the magnetizing force again, branch 2 of the hysteresis loop is obtained.

If the inducing field be given a value $-H_o$ from point SA, and then be reduced to zero, the resulting induction will first follow down curve 1 from SA to point C_1 and then will follow the minor ascending curve 3 to point C' on the induction axis B. Curve 3 cannot be retraced from C' by again applying a magnetizing force $-H_o$, but instead, curve 4 will be traced. This shows that the magnetic processes are not reversible, and that the magnetic conditions obtained from a given magnetic stimulus depend upon the magnetic prehistory.

Such hysteresis loops are obtained with a closed magnetic circuit containing only the magnetic material. However, as soon as an air gap is introduced, an effect called demagnetization occurs, which reduces the induction in the material. The shorter the length of the magnet compared with its cross section, the greater becomes the demagnetization. Materials having a high coercive force have a high resistance to demagnetization. Therefore, where magnets of relatively short lengths are to be used, it is desirable to use magnetic materials having a high coercive force.

The Magnetic Sound Carrier

A high-carbon or tungsten steel or a similar alloy has this required high coercive force.

The hysteresis loop of the material should have a relatively long straight portion to provide a high dynamic range with little distortion.

Both steel tape and steel wire are practical for recording. However, steel tape has the additional advantage that a more definite relationship between the pole pieces and the sound carrier can be maintained.

From a practical point of view, the tape must lend itself to rolling in the mill, and must be flexible for use. Since it is normally very thin, it must have a high tensile strength, and its surface must be highly polished. It should not be affected by climatic conditions and should have a high resistance against rusting. This last requirement is difficult to meet, and usually the tape surface is protected by a rust-resisting oil film.

The Pole Pieces

The pole pieces serve as the coupling link between electric action in the surrounding coil and the magnetic condition of the sound carrier. In the recording and obliterating processes, these pole pieces focus the magnetic flux in the steel tape. In the reproducing process, on the other hand, the magnetic flux from the tape has to penetrate the pole pieces to induce an electromotive force in the coil. Since the pole pieces are subjected to a varying magnetic flux, their hysteresis effect must be reduced to a minimum. This may be

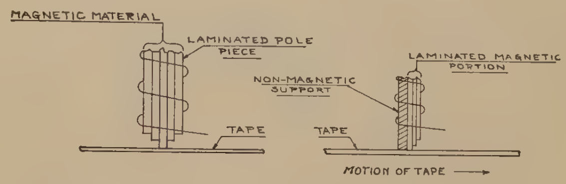


Fig. 3—Examples of possible pole-piece structure.

done in two ways: by introducing an air gap in the magnetic circuit, and by selecting a material with small hysteresis losses (low coercive force). The pole-piece tip should have the smallest possible dimension in the direction of the motion of the sound carrier. However, during the recording process, the pole piece must be able to handle without saturation a considerable magnetic induction, so as to utilize the full magnetic range of the tape.

In the reproducing process, on the other hand, the pole pieces are subjected to very weak magnetic fields, and the principal consideration is that the resulting flux lines should completely thread through these pole pieces. The reproducing pole pieces, therefore, should have the lowest possible magnetic resistance. These requirements for pole pieces are conflicting, and experience has shown that materials like mu metal, permalloy, or Allegheny metal are satisfactory compromises.

Since, during the recording process, a certain leakage flux is present, the tip of the pole piece will not be subjected to the same magnetic intensity as the center portion. Accordingly, a pole piece having a larger cross section toward the center of the coil is somewhat more effective. To avoid eddy currents, pole pieces of this kind are laminated, as shown in Fig 3.

In operation, the pole pieces are pressed against the signal carrier, and are subjected to incidental mechanical shocks. The materials which are suitable for pole pieces have a very low elastic limit, and the magnetic properties of such a material are greatly impaired when it is subjected to shock. Since the pole-piece tip has to carry the controlling flux, a mechanical shock imparted thereto greatly affects the recording and reproducing processes. It was found that this difficulty can be overcome by uniting a mechanically strong support strip of shock-absorbing material with the magnetic pole-tip laminae, into an integral pole-piece structure such as shown in Fig. 3. The nonmagnetic

support is placed ahead of the magnetic lamination, so as to take up the force of the shock imparted to it.

Longitudinal and Perpendicular Recording

During the process of manufacture of steel tape, a magnetic preferred axis is produced by the rolling action in the material in the direction in which the tape is rolled. Unless specially processed, such tape will produce, under the influence of a given field, a much stronger magnetic induction along its preferred longitudinal axis than along any other axis. Advantage can be taken of this preferred axis by displacing somewhat the planes of the recording pole pieces, and recording the signal along this preferred axis. This type of magnetic recording, shown in principle in Fig. 4, is called longitudinal recording.

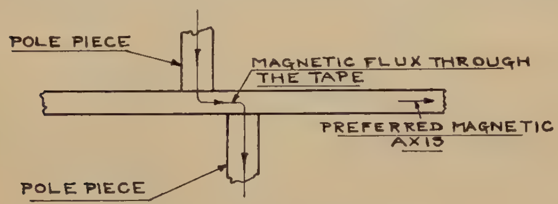


Fig. 4—Magnetization pattern in longitudinal recording.

The preferred axis of such material may be destroyed by a suitable heat treatment, and a more homogeneous magnetic material obtained. A heat-treated magnetic carrier can be magnetized equally well along any axis, and is used for perpendicular recording, in which the pole pieces are located in the same plane, and the flux is impressed perpendicularly to the axis of the signal carrier, as shown in Fig. 5.

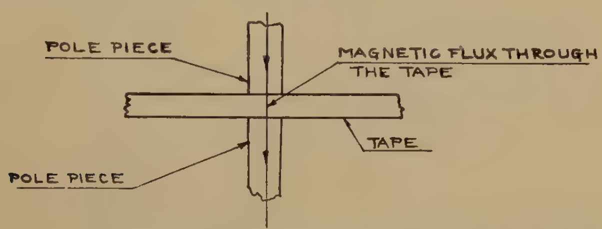


Fig. 5—Magnetization pattern in perpendicular recording.

The Obliterating Process

In order to wipe out a previous recording and put the steel tape in condition for accepting a new recording, it has to be subjected to an obliterating process. Obliteration is effected either by bringing the tape to a uniform magnetic induction, by completely saturating it, or by completely eliminating all magnetism in the tape.

The latter method can be carried out by rapidly reversing a diffuse magnetic flux, the amplitude of which decreases as the tape moves away from the center point of this flux distribution. However, the recording on such a carrier would suffer distortion on account of the curvature of the virgin branch *a* of the magnetizing loop.

The method of saturating the tape offers several advantages, and seems to be the only one used at present.

The Recording Process

As shown in Fig. 6, after the tape leaves the obliterating head, its magnetic state is represented by the residual magnetism R_1 of the magnetization loop. During the recording process, the signal current is superimposed upon a polarizing current. The resultant magnetizing force is therefore a unidirectional pulsating force (H_p+H_s). In the idealized case, recording takes place when successive tape elements are subjected to successive values of this pulsating flux. To explain this operation, one particular momentary value of recording current (represented by the magnetizing force H_m), will be analyzed. On being subjected to this momentary recording force, the magnetic condition of the tape is carried down to point B_m of the magnetization curve. Upon leaving the recording head, the magnetizing force acting on the tape element drops to zero, and the magnetic state of this tape element increases along the minor ascending loop from B_m to B_m', where B_m' now represents the final magnetic condition of the tape element. It has been found that the minor ascending loops introduce negligible nonlinearity. Since any wave form may be considered as a sequence of such points, the recording of a complete wave front may be synthesized accordingly.

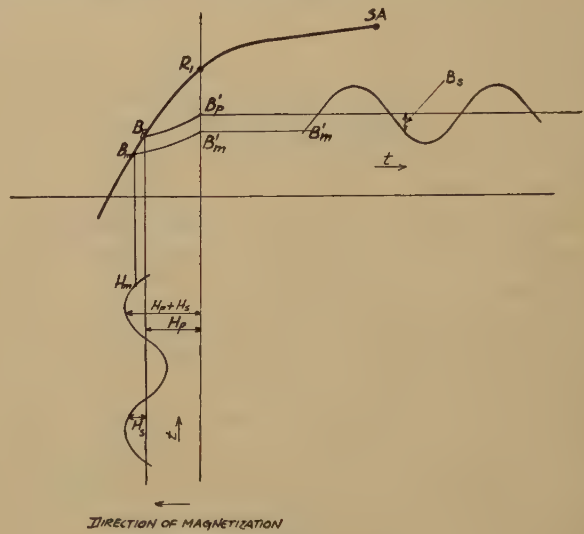


Fig. 6—The magnetization process.

The recording operation may be compared to a conventional class A vacuum-tube amplifier, with the grid-bias voltage represented by the polarizing current, and the signal voltage varying in both cases about this bias as a datum. As in the amplifier tube, this datum point should be as close to the center of the straight portion of the curve as possible for maximum undistorted power output.

Thus far the idealized case was considered of an incremental portion subjected to only one instantaneous influence of the recording flux. In the practical case, the

incremental tape portion is subjected for a certain finite length of time to the magnetizing force. This length of time is a function of the tape speed, the effective pole-piece thickness, and the magnetic focusing effect. The effect of tape speed need not be considered, since it is obvious.

In perpendicular recording, where the pole pieces are aligned in the same plane, the magnetic flux is relatively evenly distributed over the entire area of the pole-piece face. This means that the incremental portion of the tape as it passes under the face of the pole pieces receives not one instantaneous magnetic influence, but instead, receives a continuing magnetic influence all the time the incremental tape portion is in contact with the pole piece. Since the magnetic processes are not reversible, the incremental tape portions will have a final magnetic state imposed upon them by the highest magnetic stimulus they receive while trav-

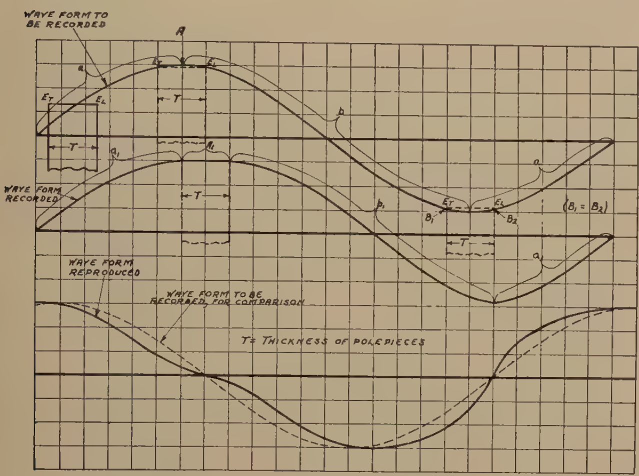


Fig. 7—Distortions generated by finite thickness of pole pieces.

eling under the pole pieces. This final induction in the tape will be left by the highest magnetizing force, irrespective of the position of the moving incremental portion of the tape under the pole pieces. This, to use an unscientific word, "smears" the recorded signal. Since it is particularly bad for higher-frequency signals, this "smear" effect represents one of the factors which limits the usuable frequency range of magnetic recording. If the pole-piece thickness approaches the order of one wavelength, no recording will take place at all. The distortions due to this smearing appear to be even harmonics. Since this type of distortion is mainly a function of the wavelength, even a small amplitude of recording will have some distortion.

To obtain a quantitative picture of the condition in the recording process, which condition is due to the finite thickness of the pole piece, it is well to consider a sinusoidal wave form and a pole piece moving relative to it, as shown in Fig. 7. On branch a of the sinusoidal wave form, the leading edge E_L is subjected to the highest magnetizing stimulus, and therefore the induction left in the tape will correspond to branch a of the wave form and will produce an induction a_1. As soon as the leading edge of E_L reaches the upper crest A of the wave form and passes it, the maximum stimulus will be shifted from the leading to the trailing edge E_T. During this period of time required for the pole piece to move along a from E_L to E_T, the recorded induction will correspond to the constant value of A, and the induction obtained is A_1. While the pole piece moves farther on along branch b of the wave form, the trailing edge E_T will now be subjected to the maximum

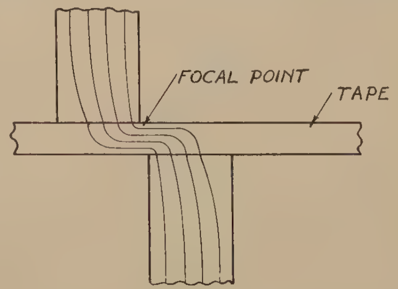


Fig. 8—Focusing of magnetic flux in longitudinal recording.

magnetic stimulus, and the magnetic induction in the tape will correspond to the shape of branch b, and will produce an induction b_1, until the trailing edge of the pole piece and the leading edge simultaneously reach values B_1 and B_2, respectively. At this instant, the induction in the tape abruptly ceases being determined by branch b of the magnetizing curve, and begins to be influenced by branch a.

The above method of analysis, of course, can also be applied to more complex wave forms, such as occur in actual recording.

In longitudinal recording, the pole pieces are offset on opposite sides of the tape to direct the flux along

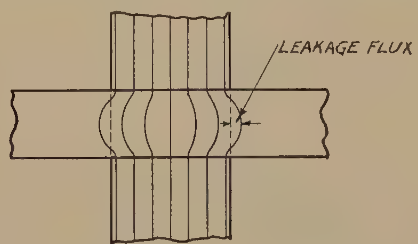


Fig. 9—Spreading of flux in perpendicular recording.

the axis of the tape. There is a shortest distance from one pole piece through the tape to the other pole piece, and since this shortest distance has the least magnetic resistance, it will have the greatest flux density. Furthermore, all of the flux lines threading between the pole pieces will pass through a cross section of the tape located between the pole pieces. In longitudinal recording, the effective magnetizing area is not so much a function of the thickness of the pole pieces as it is of the offset of the pole pieces, and the magnetizing flux can be focused in a relatively small tape length, as indicated in Fig. 8. This length corresponds to the pole-piece face area in perpendicular recording, as far as the distortions outlined in the last paragraphs are concerned.

In magnetic recording, an uncontrolled or leakage flux must be considered, as well as the main flux. In perpendicular recording, the main flux spreads somewhat as it penetrates the tape, as indicated in Fig. 9, and the spread of these flux lines beyond the thickness of the pole pieces is called the leakage flux. This leakage

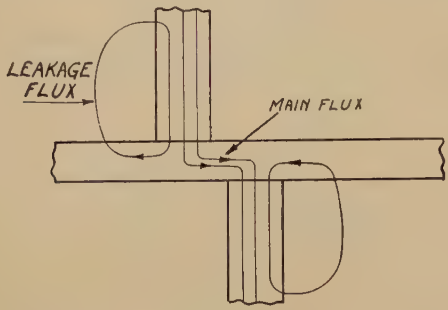


Fig. 10—Leakage flux in longitudinal recording.

flux increases the effective thickness of the pole pieces, and is reduced by employing a thin steel tape.

In longitudinal recording, the flux which threads through the tape and only one pole piece is designated as leakage flux. As shown in Fig. 10, the leakage flux has a direction opposite to that of the main flux. The leakage flux threading the tape before the tape reaches the recording pole pieces has no effect, since its direction tends to magnetize the tape in the direction in which it has already been saturated by the obliterating process. The trailing leakage flux, however, tends to superimpose a magnetizing force upon the already recorded signal, and thus interferes somewhat with this recorded signal. Proper selection of the polarizing current is a factor in the reduction of this leakage flux.

Sound-Carrier Flux Distribution

Another factor entering into the recording process is the tendency for magnets of short lengths to become

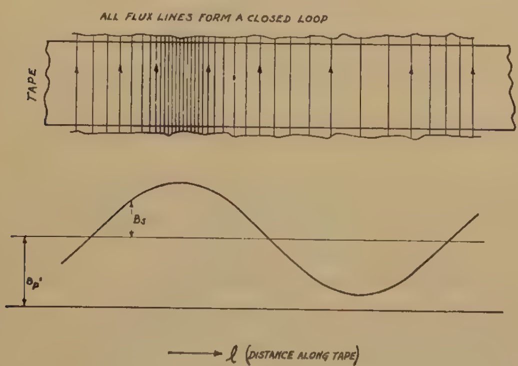


Fig. 11—Flux distribution in tape after perpendicular recording.

demagnetized. When the pole pieces with their small magnetic resistance form a part of the magnetic circuit during recording, this circuit, as far as the tape is concerned, may be considered as practically closed. When the pole pieces are removed as the recorded tape element leaves the pole pieces, this magnetic circuit is broken, and a certain amount of demagnetization takes place. In perpendicular recording, this loss is constant for all frequencies, since the length of the signal magnets are given only by the thickness of the tape. In longitudinal recording, however, the length of the magnets are proportional to the wavelengths recorded, and therefore the demagnetization will be more severe for the higher frequencies.

In Fig. 11 is shown a schematic diagram of the flux distribution in the tape after it has undergone a perpendicular recording process. All the flux lines are directed in the same way in passing through the tape and into the air, where they complete their magnetic circuit. The signal, in this case, consists only of changes of density in the magnetic pattern.

In Fig. 12 the same condition for longitudinal recording is illustrated. The longitudinal axis of the tape is a magnet of varying intensity but of the same direction. Since all magnetic lines must be continuous, the induction in the air in this case consists of the leakage flux from successive elements of this main magnet. This leakage flux is proportional to $d\phi/dl$ and, there-

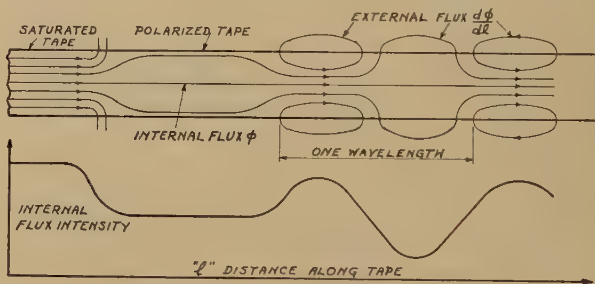


Fig. 12—Flux distribution in tape after longitudinal recording.

fore, increases with the signal amplitude and inversely with the signal wavelength. If the recording is made with the same maximum magnetizing force for increasing frequencies, the amplitude of the external flux will therefore increase 6 decibels per octave.

The Reproducing Process

In reproducing a record made with perpendicular magnetization, the external magnetic flux will thread through each pole piece, and electromotive forces will be generated in the coils surrounding each pole piece. These electromotive forces are proportional to the rate of change of the flux penetrating the pole piece. Assuming that the record has been made with constant maximum flux density independent of frequency, the voltages produced in these coils will increase 6 decibels per octave up to the point where the pole-piece width approaches the order of magnitude of the recorded wavelength.

Since in perpendicular recording, the two pole pieces are in the same plane and the flux impressed upon the tape is perpendicular to the faces of the tape, the flux lines penetrating one pole piece will also penetrate the other. This effect is independent of frequency, and so the magnetic phase relationship of the two pole pieces is always the same, independent of frequency.

Accordingly, the phase relationship between the electromotive forces generated in the two coils is fixed, and by properly connecting the two coils, it is always possible to obtain the highest generated voltage. The problem of reproducing is somewhat more complex than the above simple picture would indicate, inasmuch as the original external flux pattern of the signal is changed somewhat by the proximity of pole pieces. Fortunately it is found in practice that this does not seem to cause any distortion, and experiments show that for very long wavelengths, a 6-decibel increase per octave, as predicted by the simplified theory, actually exists.

It has been pointed out that due to the finite dimensions of the pole-piece tip, the induction recorded in the tape is not a facsimile of the recording signal. This is shown in Fig. 7. It is of interest to analyze the results of reproducing this distorted wave form with a reproducing pole piece of the same finite width. The reproduced signal is proportional to the derivative of the total flux lines threading through the pole pieces. It can be seen that the finite-thickness pole piece, when tracing this recorded wave form, will introduce a somewhat compensating distortion. This reproduced signal is also shown in Fig. 7.

In longitudinal recording, as in perpendicular recording, the voltages generated in individual coils depend mainly upon the rate of change of the external flux. It will be recalled that in longitudinal recording, the external flux density increases 6 decibels per octave. Therefore the voltage generated in each individual pole-piece coil will, under ideal conditions,

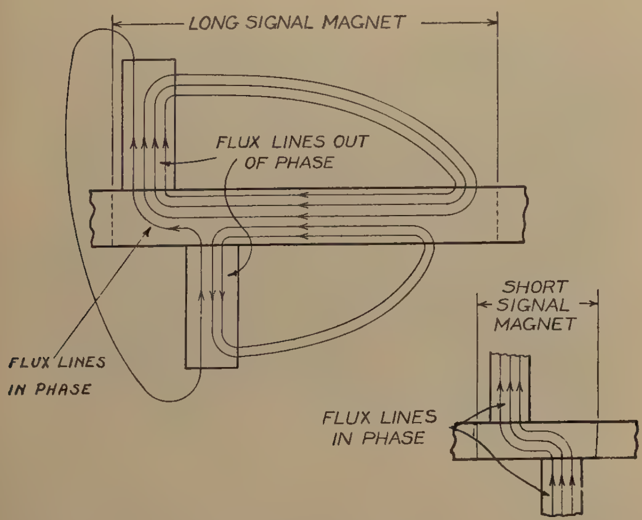


Fig. 13—Phase relationship of electromotive forces generated by a longitudinal recording.

increase 12 decibels per octave. However, certain complications must be expected, since the pole pieces are slightly offset, and the magnetic difference to which they are subjected depends upon the length of the recorded magnets, which are proportional to the recorded wavelength. Since each pole piece is subjected to a different potential of the magnetic field, the phase relationship of the electromotive forces generated in the surrounding coils will vary with the recorded wavelength. For example, as shown in Fig. 13, in reproducing a record of very low frequency formed by individual signal magnets, the potential difference between the two pole pieces is very small, and therefore

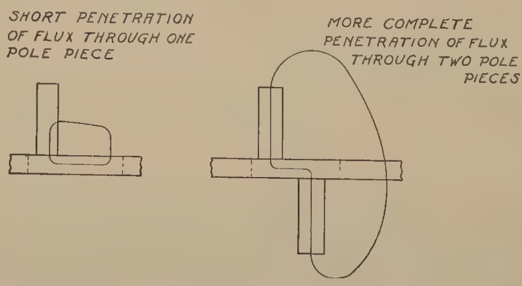


Fig. 14—Better penetration of flux through two than through one reproducing pole piece.

the induced electromotive forces of the two coils would be out of phase. On the other hand, if the recorded wavelength is so short that the magnetic-potential difference between the pole pieces is a maximum, the induced electromotive forces will be in phase. To meet these conditions, either only one coil is used in the reproduction process, or, if both coils are employed, a phase-control network is introduced to secure the best relative phase conditions of the two coils.

In practice, it has been found satisfactory to reproduce with two pole pieces, using a coil on one pole piece only.

Two pole pieces are used because, as shown in Fig. 14, the magnetic-potential difference produces a common flux which penetrates the two pole pieces further than in the case of one pole piece.

As has been pointed out above, theoretically, the electromotive forces generated in the coils should increase 12 decibels per octave for longitudinal recording. Experiments prove that this is true for very low frequencies, but at higher frequencies, two factors tend to reduce this increase. One factor is that with higher frequencies, the lengths of the recorded magnets in the tape are shorter and therefore lose more by demagnetization; the second factor is that for higher frequencies, the flux penetration through the pole pieces is somewhat reduced.

In longitudinal recording, the same considerations as for perpendicular recording can be applied to analyze the wave form reproduced by the finite thickness of the pole piece from the distorted recorded wave form. If the face of the reproducing pole-piece tip is of larger or of the same order of thickness as the length of the focal point, a similar compensating effect as that already discussed will reduce the actual distortions.

Typical Characteristics

In a typical system where suitable equalizing means are used in the recording and reproducing processes, a response frequency flat within ± 2 decibels from 60 to

10,000 cycles is obtainable. The signal-to-noise ratio is normally of the order of 40 decibels, but if particular attention is given to this point, and a selected tape is used, up to 50 decibels may be obtained.

The output level of the reproducing heads after equalization is of the order of microphone level.

Nonlinear distortion at 400 cycles is less than 2 per cent.

The Driving System

As in any other recording system, the sound carrier has to be moved with constant speed. Any flutter and speed variation will cause "wow" or distortion. It is outside the scope of this paper to discuss in detail the different drive systems which could be employed. In general, two different kinds of magnetic tape drive systems are used. The one system is based on the reel principle, requiring a rewinding of the tape after recording and before reproduction; the other principle makes use of an endless tape which moves continuously in the same direction. The continuous drive has its particular application where continuous repetition of the record is desired, where simplicity of operation is required, or where immediate reproduction of the signal is necessary.

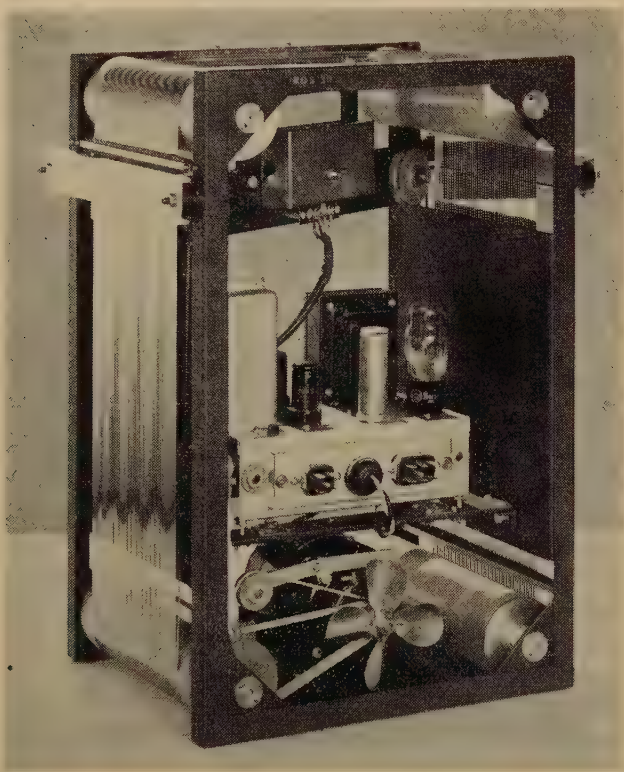


Fig. 15—Magnetic tape recorder with continuous length of tape.

An example of an endless-tape drive system is shown in Fig. 15. In this machine, the tape is helically wound on rollers at the four corners of a rectangular frame. Space within this frame is utilized by placing therein the motor, amplifier, and other associated equipment. For some applications, a synchronized timer is provided to control the recording time. Such machines have been described in the literature.

The Application of Magnetic Tape Recording

Since in magnetic recording the previous record can be obliterated simultaneously with the making of a new recording, the sound carrier can be used continuously for new recordings without any damage to the recording medium. Magnetic recording, furthermore, is not greatly affected by mechanical vibration and therefore lends itself to any application wherein a recording must be made under severe conditions of mechanical vibration. Finally, a record on a magnetic sound carrier, if not oblitereated, can be reproduced thousands of times with negligible change of quality in the reproducing process.

In the broadcasting field, the fact that the same recording medium can be used repeatedly seems to be the most important point. It immediately suggests that all recordings which have to be time-delayed for a relatively short period could be recorded on tape instead of on disk. Mechanical recording has been developed to such a high degree of perfection that there is no question in the mind of the writer that there will always be a field for it where permanent recordings are required, but as soon as only a temporary time delay of a recording is the object, disk recording seems no longer economical.

This has been the conclusion of many European broadcasting stations. Magnetic tape recording is and has been used to a great extent in England, Germany, Italy, and other countries. An article by Barrett and Tweed[1] states with regard to the use of magnetic tape recording, "In the past year, machines of this type have been on the air for reproducing program material 1957 hours." This indicates that in 1937, magnetic tape recording machines were used on the average, about $5\frac{1}{2}$ hours every day in the year. It can be assumed that these machines are now employed much more than they were a few years ago.

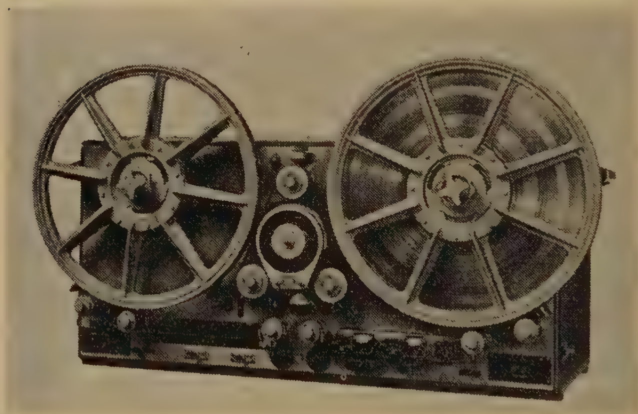


Fig. 16—Reel-type tape recorder used in Germany.

[1] H. E. Barrett and C. J. F. Tweed, *Jour. I.E.E.* (London), vol. 82, p. 286; March, 1938.

The tape recording machines which are in use in Europe were designed to record half-hour programs. The method of winding the tape is similar to the well-known reel system of a motion-picture camera. Fig. 16 shows a unit which is used at present in Germany; Fig. 17 shows a unit used in England. The English machine is somewhat larger since it makes special provisions for reduction of flutter to the very minimum. In this country units of this kind are not yet available, partly because the delay of radio programs has only recently become sufficiently extensive to justify the investment for the development of such machines.

Artificial Reverberation

This kind of delay of program material, however, is only one of the possible applications of a magnetic tape recorder in the broadcasting field. In many cases, a short time delay can be used to accomplish sound effects which may enhance the impressiveness of a program material considerably, and it is for these applications that magnetic tape recording is particularly suitable. One such application is a so-called artificial-reverberation unit.

It is a well-known fact that if sound is generated in a room by a sound source and the sound generation from this sound source stops suddenly, it takes some time until the complete sound energy in the room decays to the threshold of hearing. This is due to the fact that the original sound has been reflected back and forth between the walls of the room and has actually built up the sound energy in the space between the walls of the room. Only after a certain number of reflections is the sound sufficiently absorbed by the walls. In general, the decay of any one frequency follows an exponential characteristic.

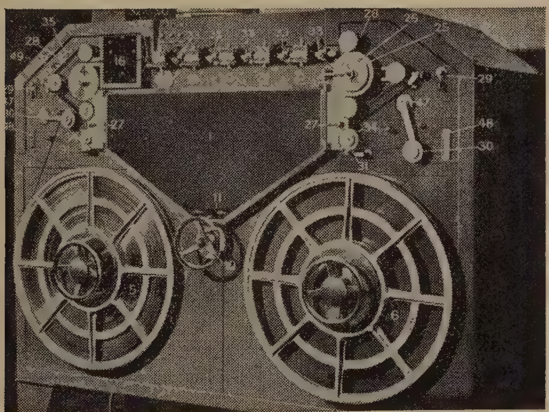


Fig. 17—Reel-type tape recorder used in England.

The purpose of an artificial reverberator is to simulate this real reverberation, or, in other words, to impress upon the listener the psychological effect of increasing the size of the room in which the program originates. In a practical unit of this kind, a recording head and a number of reproducing heads are provided. These heads are closely spaced so that only a very small time interval elapses for the sound carrier to move from one head to the next. Each of the reproducing heads is properly attenuated with respect to the preceding head, so that a signal reproduced by each of the different heads will be exponentially reduced with time when passing through the series of heads.

Fig. 18 shows a unit of this kind, built experimentally, which employs an endless magnetic tape. The signal in the tape is obliterated continually in each cycle of the tape, and the signal to be reverberated is continually recorded upon the tape. An endless-tape unit

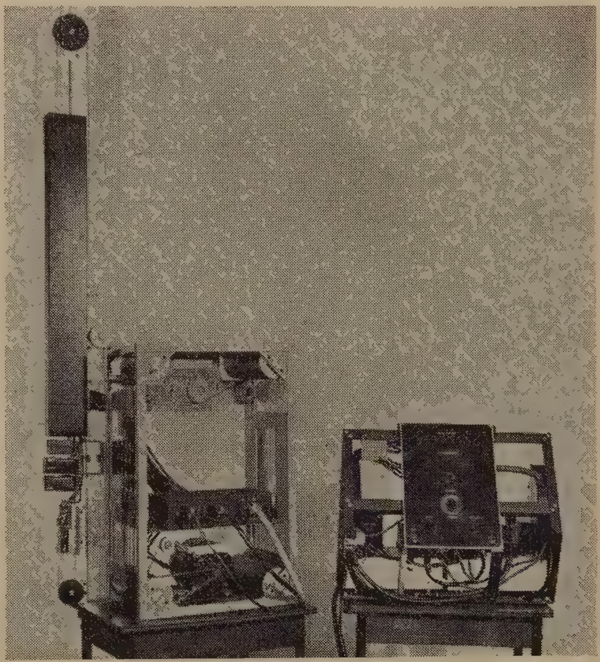


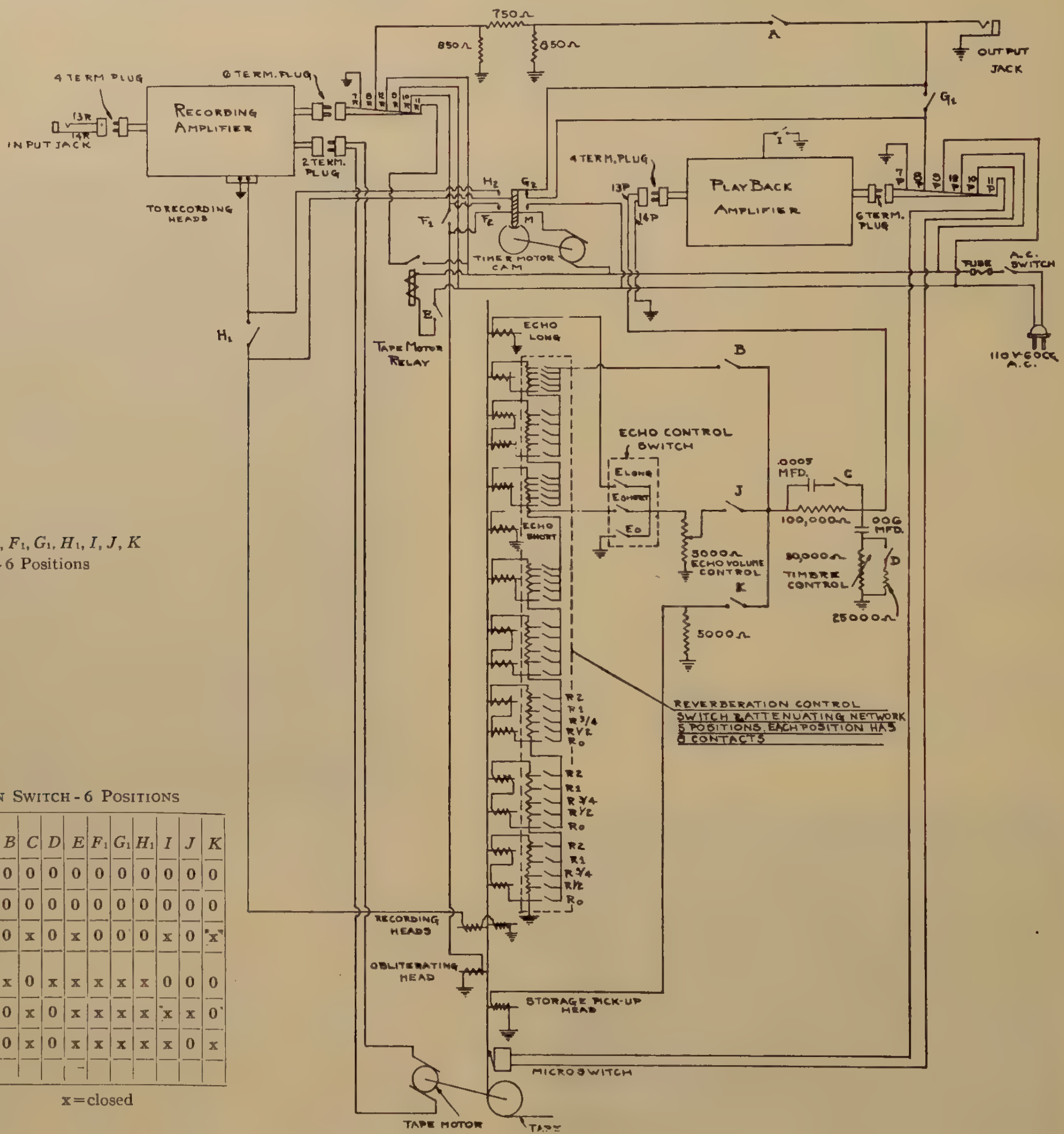
Fig. 18—Experimental artificial reverberator using magnetic tape.

has been chosen for this application since it is the most convenient, requiring no reversal of the sound carrier, and thus permitting continuous operation of the system without interruption.

Artificial reverberation is not quite the same as true reverberation. In an artificial reverberation unit of this kind, one always has to deal with distinct reflections, while in a properly treated three-dimensional room, after the first distinct echoes, a blending effect takes place, which eliminates any separations of the reverberated signals. An artificial reverberation machine can be more easily related to a room having only one pair of opposite walls acoustically alive, the other walls being completely sound-absorbing. Nevertheless, the psychological effects created by the use of an artificial-reverberation unit are so similar to those of real reverberation that a machine of this kind can certainly create a practical illusion. This is all the more true if the original sound which has to be reverberated already contains a certain amount of true reverberation which helps to provide additional blending of the different distinct signal reproductions.

Fig. 19 shows a circuit diagram of such an artificial-reverberation unit. The tape passes the obliterating and recording head and then goes through a number of playback heads. These playback heads have attenuating networks of different attenuating factors. The amount of attentuation can be changed by means of push buttons so as to give the illusion of different other heads are provided for producing echo effects, the echo being a specific kind of reverberation.

The idea of artificial reverberation is old. It has been proposed in the past to supply artificial reverberation by electrical time-delay circuits, and also by mechanical recording which permits instantaneous playback with a plurality of playback heads.



Switches $A, B, C, D, E, F_1, G_1, H_1, I, J, K$ make up main switch - 6 Positions

CHART OF MAIN SWITCH - 6 POSITIONS

	A.C. Switch	A	B	C	D	E	F_1	G_1	H_1	I	J	K
Off	0	0	0	0	0	0	0	0	0	0	0	0
Stand by	x	0	0	0	0	0	0	0	0	0	0	0
Record	x	0	0	x	0	x	0	0	0	x	0	x
Reverberation	x	x	x	0	x	x	x	x	x	0	0	0
Echo	x	x	0	x	0	x	x	x	x	x	x	0
Delay	x	0	0	x	0	x	x	x	x	x	0	x

0 = open x = closed

Fig. 19—Simplified circuit diagram of reverberator of Fig. 18.

lengths of reverberation time. All these different heads feed into a reproducing amplifier which mixes the reproduced signal with the original signal, thus adding substantially another reproducing station. These different signal combinations are then fed into the transmission line either to the transmitter or to the listening equipment.

Due to the selective frequency absorption of walls, the reverberation time varies with different frequencies. To simulate this in artificial reverberation, a variable-frequency discriminating network is provided. Certain

Considerable work has been done in the past to pro vide a simple method of producing artificial reverberation for use in broadcasting studios and the like. One of the difficulties encountered in the development of a practical sound-recording device for artificial reverberation is the fact that all practical sound-recording media have a discontinuity which makes it impossible to use them for supplying continuous reverberation effects. This is also true in the case of an endless magnetic tape, because the soldering point essential for such an endless tape introduces a discontinuity

which is very troublesome because of the large number of clicks produced in the succession of playback heads. A way was found for solving this problem, however, by using an endless tape with a relatively long playing cycle and a special control arrangement which momentarily paralyzes the playback amplifier in a special way, so that no click disturbances are observable, without in any way detracting from the continuous reverberation effect.

In the field of magnetic recording, considerable development work is being done in this country, particularly at the present time, since the commercial laboratories are becoming more familiar with the different requirements of the broadcasting and recording fields. It can therefore be expected that out of this collaboration of consumer and research, new instruments will develop which will take their places among the indispensables of the modern studio.

Acknowledgment

I am indebted in the preparation of this paper to Mr. T. E. Lynch of the Brush Development Company, who rendered assistance in the preparation and correction of the manuscript.

Bibliography

(1) V. Poulsen, United States Patents: No. 661,619, original wire recording machine. No. 873,083, wire and tape recording with perpendicular magnetization. No. 789,336, wire spool machine. No. 873,078, longitudinal magnetization recording.
(2) V. Poulsen, "The telegraphone," *Ann. der Phys.*, vol. 3, pp. 754–760; 1900.
(3) V. Poulsen, "The telegraphone," *Electrician*, vol. 46, pp. 208–210; November 30, 1900.
(4) L. Rellstab, "The telegraphone," *Electrotech. Zeit.*, vol. 22, pp. 57–59; January 17, 1901.
(5) J. H. West, "The telegraphone," *Electrotech. Zeit.*, vol. 22, pp. 181–184; February 21, 1901.
(6) K. Stille, "Electromagnetic sound recording," *Electrotech. Zeit.*, vol. 51, pp. 449–451; March 27, 1930.
(7) S. J. Begun, "Telephone recording machines for the reception of rapid telegraphy," *Telegraphen-Praxis*, no. 5, 1931.
(8) S. J. Begun, "Dictating machine in industry," *Electrotech. Zeit.*, vol. 53, pp. 204–205; March, 1932.
(9) E. Hormann, "Zur Theorie der Magnetischen Tonaufzeichnung (On the theory of magnetic sound recording)," *Elek. Nach. Tech.*, vol. 9, pp. 388–403; November, 1932.
(10) E. Meyer and E. Schüller, "Magnetic recording of sound on steel bands," *Zeit. für Tech. Phys.*, vol. 13, pp. 593–599; December, 1932.
(11) S. J. Begun, "Contribution to the theory of electromagnetic sound recording on steel tape," "Dissertation, Sonderdruck Studentenhaus Charlottenburg," 1933.
(12) Magnetic recording and reproducing—Marconi-Stille apparatus, *Wireless World*, vol. 34, pp. 8–10; January 5, 1934.
(13) N. M. Rust, "Marconi-Stille recording and reproducing equipment," *Marconi Rev.*, no. 46, pp. 1–11; January–February, 1934.
(14) H. A. Frederick, "Recording and reproducing sound," *Rev. Sci. Inst.*, vol. 5, pp. 177–182; May, 1934.
(15) H. J. Von Braunmühl, "Magnetic sound recording in broadcasting service," *Funktech. Monatshefte*, no. 12, pp. 483–486; December, 1934.
(16) R. F. Mallina, "Mirror for voice," *Bell Lab. Rec.*, vol. 13, pp. 200–202; March, 1935.
(17) T. Volk, "A.E.G. magnetophone," *A.E.G. Mitteilungen*, no. 9, pp. 299–301; September, 1935.
(18) T. Volk, "Magnetophone—a new sound recording apparatus," *Filmtech.*, vol. 11, pp. 229–231; October, 1935.
(19) W. H. Hansen, "The Magnetophone," *Elektrotech. Zeit.*, vol. 56, p. 1232; November 7, 1935.
(20) E. Schuller, "Magnetic sound recording," *Elektrotech. Zeit.*, vol. 56, pp. 1219–1221; November 7, 1935.
(21) H. E. Hamilton, "The blattnerphone—its operation and use," *Elect. Digest*, p. 347; December, 1935.
(22) W. F. Schrage, "Sound recording on magnetic materials," *Radio Craft*, vol. 7, pp. 537, 562; March, 1936.
(23) S. J. Begun, "The new Lorenz steel tone tape machine," *Lorenz Berichte*, p. 3; January, 1936.
(24) S. J. Begun, "The new Lorenz steel tone tape machine," *Elec. Comm.*, vol. 15, pp. 62–69; July, 1936.
(25) S. J. Begun, "Recent developments in magnetic sound recording," *Jour. Soc. Mot. Pic. Eng.*, vol. 28, pp. 464–472; May, 1937.
(26) S. J. Begun, "Magnetic recording-reproducing machine for objective speech study," *Jour. Soc. Mot. Pic. Eng.*, vol. 29, pp. 216–218; August, 1937.
(27) A. E. Barrett and C. J. F. Tweed, "Some aspects of magnetic recording and its application to broadcasting," *J.I.E.E.* (London), *Proc. Wireless Sec.*, vol. 82, pp. 265–288; March, 1938.
(28) S. J. Begun and S. K. Wolf, "On synthetic reverberation," *Communications*, vol. 18, pp. 8–9; August, 1938.
(29) S. J. Begun, "Magnetic recording," *Electronics*, vol. 11, pp. 30–32; September, 1938.
(30) C. N. Hickman, "Sound recording on magnetic tape," *Bell Sys. Tech. Jour.*, vol. 16, pp. 165–177; April, 1937.

Progress in the Development of Instruments for Measuring Radio Noise*

CHARLES M. BURRILL†, MEMBER, I.R.E.

Summary—*Following a brief discussion of the nature of radio noise, the basic problems of radio-noise measurement are outlined. The need for standardization of method is emphasized, and a review is given of activities both here and abroad directed to that end. For comparison, the present status of standardization in the related field of acoustic-noise measurement is also discussed. The objective of "readings proportional to annoyance" is stated, and the need for subjective or listening tests to measure the degree of attainment of that objective is set forth. The paper concludes with a chronological account of the different instruments which have been used to measure radio noise, with descriptions of the fundamental characteristics of each. Particular attention is given to the new quasi-peak type of radio-noise meter with logarithmic scale recently standardized by the Joint Coordination Committee on Radio Reception, of the Edison Electric Institute, the National Electrical Manufacturers Association, and the Radio Manufacturers Association.*

* Decimal classification: R270. Original manuscript received by the Institute, April 2, 1941. Presented in part, Rochester Fall Meeting, November 15, 1939. Presented, New York Meeting, May 7, 1941. † RCA Manufacturing Co., Inc., Camden, N. J.

FROM THE standpoint of language the term "radio noise" may be objected to on several grounds. Acoustical engineers are likely to object to it on the ground that radio noise is not noise at all. It has been claimed that the term confuses many. In particular, specialists in acoustic noise are dismayed when "radio noise" is referred to simply as noise, as has frequently been done. However, offsetting these reasonable objections, radio engineers find the term very useful and find no better alternative in the language. Under these circumstances, it seems desirable to define here "radio noise," and to make a plea for scrupulous care in the proper use of the term without abbreviation.

Any electrical disturbance which excites or is capable of exciting a radio receiver in such a way as to produce acoustic noise is called radio noise. Radio noise is electrical; it is admittedly not noise but something capable of producing noise. The intervention of a receiver for converting radio-frequency energy into sound is implied as an essential part of the definition.

If an electrical disturbance produces acoustic noise through the intervention of a receiver for converting audio-frequency electrical energy into sound, it may, by an analogous definition, be termed audio noise. However, such disturbances have generally been called simply noise. Just as in the case of radio noise, it is desirable to be more precise.

In facsimile and television, receivers are used for converting electrical energy into visual images. These receivers are subject to interference from extraneous electrical disturbances, just as sound receivers are. This visual interference has frequently been called noise, by analogy, although this is probably too great a stretching of the word's basic meaning. A new generic term is needed.

It should be stated that an electrical disturbance may simultaneously interfere with a radio sound receiver, a radio picture receiver, and a voice-frequency telephone circuit. It is then at the same time radio noise, radio-visual interference, and audio noise. For this reason a measurement of radio noise may serve to some extent as a measurement of radio-visual interference or of audio noise. For example, measurements of disturbances as radio noise have so far been considered satisfactory for the study of visual interferences in television.

The present paper will be limited to the consideration of radio noise and audio noise which results therefrom through the process of detection. Visual interference is a subject by itself, as yet practically unexplored, which must be left for treatment elsewhere.

Obviously, in the case of anything so generally described as an electrical disturbance, there are many characteristics which might be measured. Thus the most difficult problem in radio-noise measurement is to select, from the many types of measurements which might be made, the ones which are most significant for the purposes desired. It is easy to obtain numerical measures of radio noise; the problem is in the interpretation of these values after they are obtained.

Since the ultimate interest in the radio noise resides in the acoustic noise which it is capable of producing, the fundamental evaluation of any radio-noise measurement must be in terms of this acoustic noise. Thus the art of radio-noise measurement depends directly on the art of acoustic-noise measurement. Unfortunately, however, the two arts so far have developed practically independently of each other. In some respects radio-noise measurements have outstripped acoustic-noise measurements where logically they should have followed. This has been because radio engineers have been impatient for results and have been willing to proceed even in the absence of a clearly defined theoretical basis. Such pioneering is highly commendable in a case like this where a theoretical basis is so hard to find. There is, however, grave danger of wasted effort if such empirical progress is not followed as soon as possible by fundamental theory. There must also be a willingness to redirect empirical effort frequently to conform to new improvements in theory, else the development may fly off on a tangent.

The evaluation of a radio-noise measurement consists in correlating it with a satisfactory acoustic-noise measurement; that is, one which has been already adequately evaluated with respect to actual listening. The best type of radio-noise measurement is the one which gives the most satisfactory correlation.

However, in actual practice, many types of radio-noise measurement have been used, no one of which has as yet been scientifically evaluated by such a correlation. This has been because satisfactory acoustic-noise measurement methods have not been available or have been exceedingly cumbersome and expensive in use. The justification for each type of radio-noise measurement has been the fact that it has proved valuable in use. The relative values of the different methods, however, can be estimated only approximately from the opinions of the users of the different methods.

Much dissatisfaction with radio-noise measurements has resulted from the instinctive expectation of many people that the results of different types of measurement of the same radio noise should agree numerically. The foregoing discussion should make it clear that, unfortunately, this cannot reasonably be expected, except, as an approximation by accident or design, over a very limited range of circumstances. Each method measures a different entity, a different characteristic of the very complex electrical disturbance comprising the radio noise.

An encouragement thus to err in expecting numerical agreement is the general practice of expressing the results in voltage units. Of course the measurements may all be referred to a quantity having the dimensions of a voltage, but a different voltage is involved in each case, and in no case is this voltage present in the original electrical disturbance or radio noise. It would help toward a clear understanding of this situation if the results of radio-noise measurements were expressed as a certain number of radio-noise units in accordance with a specified test procedure. (This could be suitably abbreviated or symbolized for convenience.) Compare, for example, the standard specification of hardness (an empirical measurement) as so many units in accordance with the Rockwell C scale.

Instead of worrying about lack of numerical agreement between different types of measurement, workers should direct their attention toward adequate specification of all the requirements for each type of

measurement, so that all measurements of a given type, made with different instruments in different places, will be truly comparable. This is the first great problem—the standardization of method.

Much progress has been made in this direction and a number of organizations are active in the work, of which the following is a partial list.

In the United States:

The Joint Coordination Committee on Radio Reception of the Edison Electric Institute, the National Electrical Manufacturers Association, and the Radio Manufacturers Association.

Radio Manufacturers Association, Committee on Testing and Measuring Instruments.

Radio Manufacturers Association, Subcommittee on Television Interference.

Joint Radio Manufacturers Association-Society of Automotive Engineers Committee on Ignition Interference.

In Canada:

Dominion Department of Transport.

Canadian Engineering Standards Association.

In England:

British Standards Institution.

Post Office Department.

And in Europe generally (at least until the outbreak of hostilities):

International Special Committee on Radio Interference (Comité International Species des Perburbations Radiophoniques—C.I.S.P.R.) of the International Electrotechnical Commission.

Unfortunately there has been practically no joint action between groups in Europe and in America. There is not even a direct procedure for the interchange of data between the groups in the United States and those in Europe. As a result, the practice of radio-noise measurement has developed in this country, until recently, almost independently of similar activities elsewhere. One of the objects of the present paper is to call attention to some of this other work and to relate it to the work here.

The second great problem in radio-noise measurement is the selection of the method or methods which yield the most significant results in practical use, and the finding of changes in these methods to increase their usefulness. It is perhaps too much to expect that a single method will be found best for all types of investigations. However, it should be possible by skillful development to reduce to a few the number of methods of radio-noise measurement required in the great majority of cases. In a way, the lack of world-wide standardization is fortunate, in that it has provided us with a number of methods ready for comparison and evaluation. The sponsors of each one believe that theirs is the best they know how to devise for the field of use they have in view, but they have little to back up this faith except general experience and opinion. Probably the most enthusiastic of them would not consider his method ultimate or optimum. Selection by "the survival of the fittest," that is, evolution, may be effective in the long run, but certainly it is an exceedingly slow process. If we want real progress in radio-noise measurement, we must have thoroughly scientific and fundamental evaluations of standard and variant methods through properly controlled scientific correlations with actual listening experience. Without this, standardization will be a bar to further progress, for new methods which may prove better will then no longer be tried.

Before proceeding with an approximately chronological account of the various instruments and methods which have been used for measuring radio noise, it seems desirable to say something regarding the present status of acoustic-noise measurement.

Beginning in 1932, a sectional committee of the American Standards Association, sponsored by the Acoustical Society of America, took up the study of acoustic-noise measurement. As a result, tentative standards[1] were formally adopted on February 17, 1936, which specify a basic procedure for determining subjectively, i.e., by listening tests, the equivalent loudness of a sound in terms of the physical or objective intensity of a 1000-cycle pure tone judged to be equally loud. Substantially the same procedure was adopted in England at about the same time.[2] In Germany the same method is also used with only minor modifications. The unit of equivalent loudness thus subjectively determined has usually been called the "phon" to distinguish it from the decibel used as an objective unit to measure the intensity of the comparison tone.

In order to make this procedure serve as an accurate primary or reference standard, as intended, it has been necessary to specify very closely the conditions of listening. Unfortunately, the necessary conditions are, in general, impractical of attainment except in carefully controlled laboratory experiments which are at best very time-consuming and expensive. Other methods are required for general engineering use in the laboratory and particularly in the field. Such secondary methods should, however, be compared and standardized with the primary standard procedure.

In recognition of this need for a secondary method of measuring the loudness of acoustic noise, standards for a sound-level meter for this purpose were also adopted by the American Standards Association on February 17, 1936.[3] These standards have been very favorably received, and instruments made in accordance with them have been much used. The values indicated are called sound levels in decibels and are intended to be approximately equal to the loudness levels in phons. The instrument consists of an amplifier

[1] "American Tentative Standards for Noise Measurement, Z24.2—1936," American Standards Association, New York, 1936.

[2] "British Standard Glossary of Acoustical Terms and Definitions," B.S.S. No. 661—1936, British Standards Institution, London, 1936.

[3] "American Tentative Standards for Sound Level Meters, Z24.3—1936," American Standards Association, New York, 1936.

and root-mean-square output-voltage indicating meter, used with a microphone having as little directivity as is convenient. The frequency response is adjusted so that, for pure tones of a specified loudness, the indications should be exactly equal to the loudness levels in phons. However, the loudness contributions of the separate components of a complex sound are not combined by the ear in the same way that the root-mean-square indicating meter combines their energies, nor in fact in any simple manner. Thus for complex sounds, the indications of this instrument, or sound levels, are only approximate measures of loudness.

B. G. Churcher and A. J. King,[4] in a comprehensive paper, have ably discussed the technique of the standard subjective loudness measurement and the accuracy and reproducibility to be expected of it. They have also described comparisons with it of a number of secondary methods, including the use of an instrument which the British call an objective noise meter, similar to the American Standard sound-level meter. In this case they found, as was to be expected, increasing divergence in indications as the type of sound tested was changed from simple to complex. With an impulsive noise containing a very large number of Fourier components, produced by a 50-cycle saw-tooth oscillator, the objective noise meter read 23 decibels too low.

Another type of objective noise meter, developed expressly for the measurement of impulsive noises such as those produced by gasoline engines, has been described by A. H. Davis[5] of the National Physical Laboratory (England). This instrument uses a quasi-peak-reading output indicator very similar to those to be described more in detail later in connection with radio-noise meters. Its readings correspond well to subjectively determined loudness levels, for both steady and impulsive noises, as is shown in the paper cited.

The desired objectives in acoustic-noise measurement and the extent to which they are realized in American Standard sound-level meters are discussed in a recent paper by J. M. Barstow.[6] He compares the advantages of average, root-mean-square, and quasi-peak indicators, and concludes that the root-mean-square indication is preferable for steady and relatively steady noises. Large errors with impulsive noises are admitted but no solution is offered.

It is worth noticing, before passing on to radio-noise measurements, that all the measurements of acoustic noise referred to have been measurements of loudness. Now what is really of interest in most cases in connection with a noise is its vexatiousness, which may or may not correspond to its loudness. The annoyance caused by a noise depends on what one wishes to listen to in the presence of the noise. If what one wishes is silence, then perhaps loudness is the proper measure of annoyance. If, however, one wishes to hear music, for example, the case may be somewhat different, because of masking effects in the ear. However, for the present we shall do well to be satisfied to measure the loudness of noise.

The early investigations of radio noise were purely qualitative. A sensitive portable radio receiver, usually with a loop or directional antenna, was used chiefly to locate the source of the radio noise without attempting a numerical measure of its intensity. Such an instrument, which may be called a radio-noise locater, should be distinguished from the quantitative instrument or radio-noise meter.

The first radio-noise meters were simply radio receivers with an output indicating meter substituted for the headphones or loud speaker. The importance of the law of response of the indicating instrument to varied wave forms in determining the measured magnitudes was not realized, and root-mean-square or average-reading voltmeters and vacuum-tube voltmeters with various wave-form laws were used almost indiscriminately.

The first attempt at standardization of radio-noise measurement methods was by the Joint Coordination Committee on Radio Reception of the Edison Electric Institute, the National Electrical Manufacturers Association, and the Radio Manufacturers Association, organized for that purpose among others. In its report issued in January, 1933,[7] standard selectivity and fidelity curves for a radio-noise meter were given, in recognition of the importance of these characteristics in determining the magnitude of the indications to be obtained. The transient characteristics of the indicating instrument were specified (critical damping and an undamped period of $\frac{1}{2}$ second), but the characteristics of the two rectifiers, namely, the detector for converting the radio noise into audio noise and the rectifier for converting the audio noise into the direct current impressed on the indicating meter, were not limited in any way. Further, it was stated that measurements might be made either with or without the presence of a radio carrier, although the characteristics of detectors are almost always greatly changed by the presence of a carrier. However, the importance of specifying these rectification characteristics has not been appreciated until quite recently.

Following the issuance of these specifications, the General Electric Company manufactured a radio-noise meter in accordance with them which has been described by C. R. Barhydt.[8] This instrument, which is shown in Fig. 1, used a triode as a plate-curvature type

[4] B. G. Churcher and A. J. King, "The performance of noise meters in terms of the primary standard," *Jour. I.E.E.* (London), vol. 81, pp. 57–90; July, 1937.

[5] A. H. Davis, "An objective noise-meter for the measurement of moderate and loud, steady and impulsive noises," *Jour. I.E.E.* (London), vol. 83, pp. 249–260; August, 1938.

[6] J. M. Barstow, "Sound measurement objectives and sound level meter performance," *Jour. Acous. Soc. Amer.*, vol. 12, pp. 150–166; July, 1940.

[7] "Radio Noise Measuring Set," National Electric Light Association, Publication No. 32, January, 1933.

[8] C. R. Barhydt, "Radio noise meter and its application," *Gen. Elec. Rev.*, vol. 36, pp. 201–205; April, 1933.

of detector, and a bridge type of copper-oxide rectifier to drive the direct-current indicating meter. Measurements were made by the substitution method using a standard radio-noise generator and a calibrated attenuator. A saw-tooth oscillator of the relaxation type, operating at 120 cycles per second, served as the comparison radio-noise source. This arrangement is advantageous in that the influence of the rectifier and output-indicating-meter characteristics on the measurements is partially eliminated. The extent of this elimination depends on how closely the wave forms of the actual radio noise and the comparison radio noise are alike. The comparison method, however, has the serious disadvantage of not being direct-reading. This is particularly troublesome because of the highly variable character of most radio noises. The comparison wave form was fortunately chosen, for it is typical of a large class of radio noises with a repetition rate of 120 cycles per second produced by 60-cycle power-operated devices.

In December, 1935, the Joint Coordination Committee issued another report[9] containing specifications for a measuring instrument substantially as in the previous report, together with much more detailed instructions as to its use.

Meanwhile the author as well as others had found a radio-noise meter of this type, giving indications approximately proportional to the average value of the output wave, very unsatisfactory for the study of impulsive radio noise such as is produced by the ignition systems of automobiles and the like. The average value of such a noise wave is far too low to serve as a measure of its annoyance-producing power. It appeared that a measurement of the peak output would in this case correlate better with the results of listening tests.

The author, therefore, developed for his own use a new peak-reading audio-noise meter employing a basic circuit previously worked out by his associate, A. R. Morgan, for monitoring in sound recording. This circuit, which is shown in Fig. 2, is a modification of one originally described by F. V. Hunt.[10] The device is an audio-frequency voltmeter which, when connected across the output of any radio receiver having the properly standardized characteristics, makes with it a radio-noise meter. Fig. 3 shows, at the right, the appearance of this instrument. The high-frequency field-intensity meter with which it was designed to be used is shown at the left.

This noise meter has a logarithmic scale and is direct-

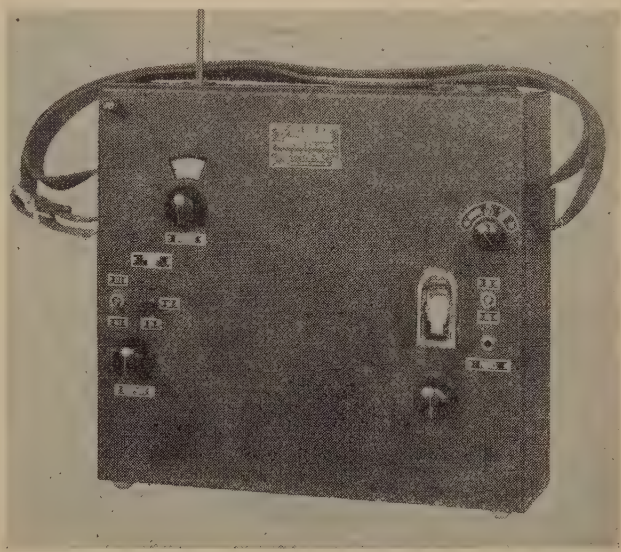


Fig. 1—General Electric radio-noise meter.

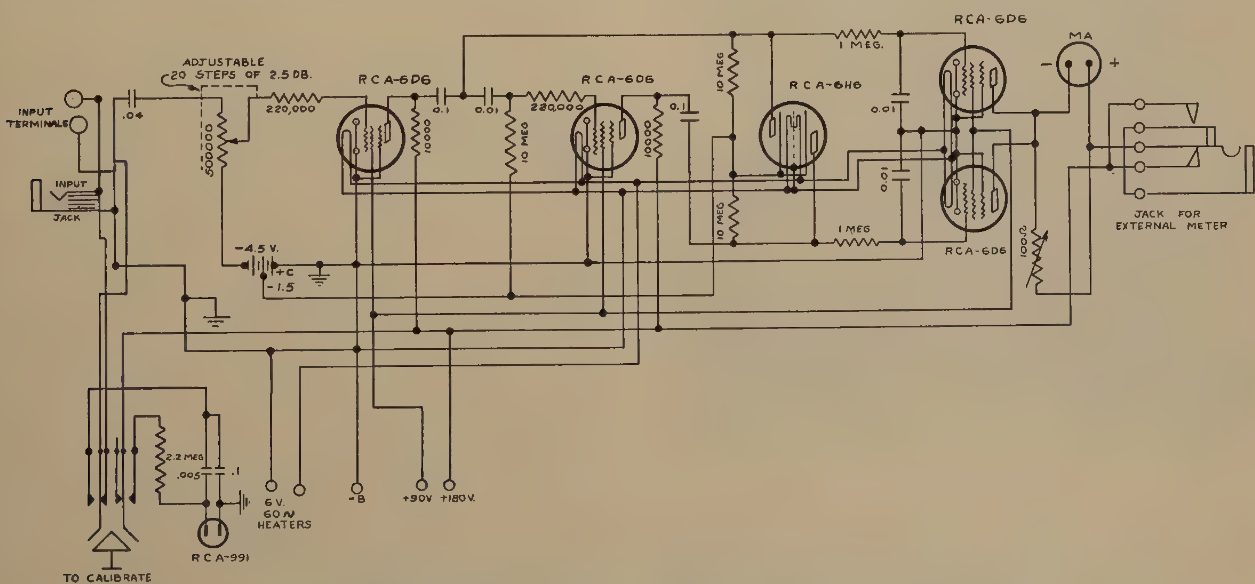


Fig. 2—Schematic circuit diagram of the RCA TMV-141A electrical-noise meter.

[9] "Methods of Measuring Radio Noise," Edison Electric Institute, Publication No. C9, National Electrical Manufacturers Association, Publication No. 102, Radio Manufacturers Association, Engineering Bulletin No. 13, December, 1935.

[10] F. V. Hunt, "A vacuum-tube voltmeter with logarithmic response," *Rev. Sci. Instr.*, vol. 4, pp. 672–675; December, 1933.

reading over a range of 100 to 1, or 40 decibels. The logarithmic scale is obtained by controlled plate-circuit distortion of wave form in two cascaded amplifier stages. The rectification is accomplished in a double-diode circuit arranged to have a short time constant effective for increasing response and a long time con-

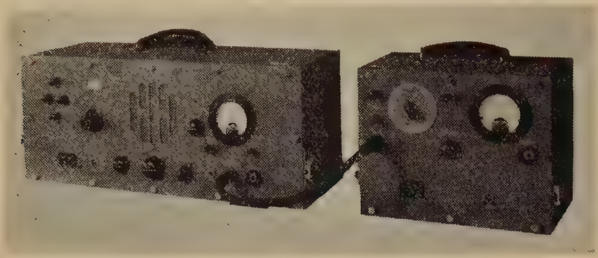


Fig. 3—RCA TMV-139A high-frequency field-intensity meter (left) and RCA TMV-141A electrical-noise meter (right).

stant effective for decreasing response. This results in an indication of the peak value unless the repetition rate of the peaks is quite slow. The calculated time constants are 10 and 1000 milliseconds, respectively. The transient-response characteristics of the instrument have been found suitable, and reproducible enough in several samples, by the test of actual use.

The first extended use of this instrument was in a survey of radio-noise levels at 41 megacycles in Manhattan streets, made from an automobile in April,

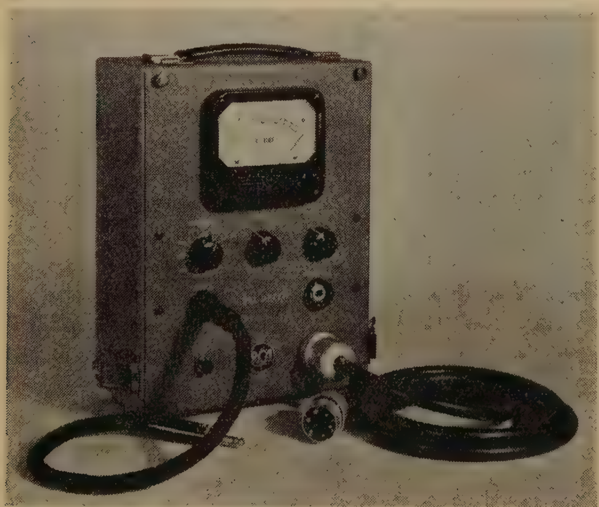


Fig. 4—RCA Type 302A noise meter.

1935. Since then it has been used in quite a variety of investigations by many of the author's associates, who have been particularly pleased with its easy readability on rapidly fluctuating radio noises. Its wide-range logarithmic scale, and the "floating" behavior of its indicating meter needle due to the quick-response—slow-recovery characteristics of the rectifier circuits driving it, have resulted in this desirable feature. Those who have used it have also been pleased with the correlation of its readings to listening experience as compared with the correlation of results with the older "average" reading instruments. No detailed account of this instrument or its use has been published.

This instrument was superseded in 1938 by the noise meter shown in Fig. 4, which was intended to have the same performance characteristics. Its logarithmic scale is obtained by the automatic-volume-control or feedback method, which requires less battery power than the "overloading" type of circuit and therefore is much more attractive for portable field use.

Under the initiative of the Joint Radio Manufacturers Association—Society of Automotive Engineers Committee on Ignition Interference, the engineering division of the Radio Manufacturers Association in December, 1935, commissioned L. C. F. Horle to develop apparatus especially suited to the investigation of automobile-ignition interference. The radio-

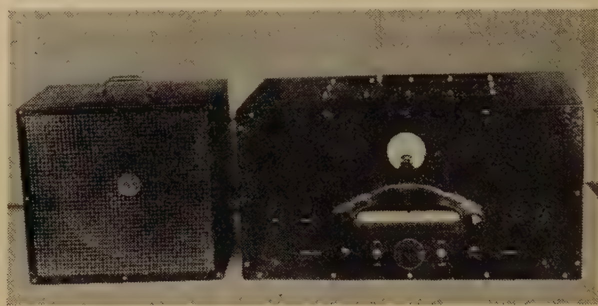


Fig. 5—Radio-noise meter developed by L. C. F. Horle for the engineering department of the Radio Manufacturers Association.

noise meter which resulted,[11] shown in Fig. 5, differed from all previous instruments in that the peak value of the noise wave was measured at the output of the intermediate-frequency amplifier instead of at the output of the audio amplifier. That is, the radio noise was measured without first converting it into audio noise. The peak measurement was made by the slideback method using a negatively biased diode detector or rectifier. The audio amplifier and loud speaker served to indicate audibly, as the gain of the intermediate-frequency amplifier was adjusted, the point where the noise peaks were just able to override the diode bias. M. G. Crosby and his associates have used a similar slideback method for measuring audio noise.

There are two disadvantages of the slideback type of peak measurement. First, it is not direct-reading, and second, personal judgment plays too great a part in determining the indication, particularly in measuring rapidly varying radio noises. However, the apparatus required is simple, and, with some care and experience, values can be obtained which check satisfactorily those obtained by the rapid-response—slow-recovery type of peak measurement.

Meanwhile the International Special Committee on Radio Interference had begun its work in Europe, with one of its objectives, "definition of a low-frequency noise indicator giving indications corresponding,

[11] L. C. F. Horle, "The development of the radio noise meter," Radio Manufacturers Association Bulletin No. 16, 1936.

whatever the nature of the noise, to the unpleasant noises due to interference coming from the loud speaker."[12] They recommended in April, 1935,[12] the use in a radio-noise meter of a vacuum-tube voltmeter of the rapid-response—slow-recovery type at the output of the audio amplifier. The following characteristics were specified for this indicator:

Time constant for the charging of the condenser	1 millisecond
Time constant for the discharging of the condenser	160 milliseconds
Time constant of the indicating meter needle	160 milliseconds
Linear scale, ratio	1 to 10

Under the leadership of the Central Electrical Laboratory of the Belgian National Committee of the International Electrotechnical Commission, the development of such an instrument was carried forward, and several models were built and tested.[13] In England, similar instruments were later designed and constructed by the General Post Office and by the Electrical Research Association.[14] The basic features of these instruments have been incorporated in a British Standard.[15]

In some of these instruments and particularly in the later ones, the rapid-response—slow-recovery type of tube voltmeter has been connected directly to the output of the intermediate-frequency amplifier, instead of to the output of an audio amplifier, as was originally contemplated by International Special Committee on Radio Interference. The two methods have been shown to be both theoretically and practically equivalent, provided, when an audio amplifier is used, it does not limit the over-all frequency band to less than that passed by the preceding radio-frequency stages, and that the detector preceding it is linear. Since practical detectors are linear to a close approximation, only when a carrier of large amplitude relative to that of the noise is present, the equivalence is close only when the audio measurement is made with a strong carrier present in the detector. Since, therefore, the radio measurement eliminates the audio amplifier, the need for a carrier, and any uncertainty relative to detector characteristics, it has come to be preferred. A further advantage in the radio measurement is that an unmodulated oscillator may be used for calibration, whereas, for the audio measurement, a modulated oscillator is required and any inaccuracy in measuring the degree of modulation is a further source of error. Therefore the later International Special Committee on Radio Interference specifications and British Standard Specification, No. 727-1937, have placed the rapid-response—slow-recovery tube voltmeter at the output of the intermediate-frequency amplifier. These specifications are outstanding in the care with which all the characteristics involving the standardization of the measurement have been defined, including requirements to prevent nonlinearity from influencing the measurement of very peaked disturbances.

In 1938 there was placed on the market in the United States a radio-noise meter designed in general conformity with the International Special Committee on Radio Interference specifications. This instrument has been described briefly by its designer, C. J. Franks, in a paper which also includes a general discussion of the principles of radio-noise measurement.[16] This radio-noise meter differed from the European instruments in that a logarithmic scale was provided by means of an automatic-volume-control circuit similar to that customarily used in radio broadcast receivers. This made the instrument direct reading over a range of 60 decibels or 1000:1 instead of 20 decibels or 10:1 as for the previous linear-scale instruments, a characteristic most advantageous in the measurement of rapidly changing radio noises. Another new feature was the provision within the instrument of means for a secondary or transfer calibration using shot noise from a saturated diode as a standard noise source. The noise output of the diode is assumed to be proportional to the average diode current, which is measured and adjusted to a predetermined value in calibration. This procedure, besides being convenient, is advantageous in correcting for small changes in selectivity of the instrument. The correction is exact for radio noise with the same wave form as that of the calibrating shot noise, approximate for other radio-noise wave forms.

The Subcommittee on Instruments and Methods of Measurement of the Joint Coordination Committee began in July, 1938, the development of new specifications for a radio-noise meter and the consideration of improved methods of measurement, with a view to publishing a report to supersede their 1935 report.[9] Three active members of that committee have given an informal account of that work in a joint paper presented before the American Institute of Electrical Engineers at San Francisco in June, 1939.[17] The formal report of the Joint Coordination Committee was published in February, 1940.[18]

The radio-noise meter specifications adopted provide for an instrument having a logarithmic scale, direct reading over a range of 100:1, or 40 decibels, in

[12] International Special Committee on Radio Interference (C.I.S.P.R.) Report R.I.3, April, 1935.

[13] International Special Committee on Radio Interference, Reports R.I. (Belgium) 1, November, 1937 and R.I. (Belgium) 2, December, 1937.

[14] A. J. Gill and S. Whitehead, "Electrical interference with radio reception," *Jour. I.E.E.* (London), vol. 83, pp. 345–394; September, 1938

[15] "The Characteristics and Performance of Apparatus for the Measurement of Radio Interference," B.S.S. No. 727—1937, British Standards Institution, London, 1937.

[16] C. J. Franks, "The measurement of radio noise interference," *R.M.A. Eng.*, vol. 3, pp. 7–10; November, 1938.

[17] C. V. Aggers, D. E. Foster, and C. S. Young: "Instruments and methods of measuring radio noise," *Elec. Eng.*, vol. 59, pp. 178–192; March, 1940.

[18] Methods of Measuring Radio Noise," Edison Electric Institute, Publication No. G9, National Electrical Manufacturers Association Publication No. 107, Radio Manufacturers Association Engineering Bulletin No. 32, February, 1940.

which a tube voltmeter of the rapid-response—slow-recovery type measures the output of the intermediate-frequency amplifier. An internal calibrator is specified with the recommendation that it be of the shot-noise type.

The time constants adopted for the rapid-response—slow-recovery circuit are 10 milliseconds and 600 milliseconds instead of the 1- and 160-millisecond time constants recommended by The International Special Committee on Radio Interference. This deviation in practice was decided upon after much deliberation, because the following considerations appeared to have greater weight than the natural desire to adopt a previous standard.

(1) A 1-millisecond charging time for the capacitor of a diode rectifier is difficult to obtain in practice without having the variable resistance of the diode make the actual time constant vary with the magnitude of the impressed voltage, or without using a very small capacitor such that an impractically large discharge resistor must be used to obtain the desired discharge time constant.

Fig. 6—RCA Type 312A radio-noise meter showing control panel.

(2) With a discharge time constant of 160 milliseconds, the transient characteristics of any practical indicating instrument which may be used will play an important part in fixing the over-all response of the instrument, and hence these characteristics must be closely specified. On the other hand, with a discharge time constant of 600 milliseconds, an indicating instrument may be used which is fast enough in response so that its characteristics are relatively unimportant and hence need only approximate specification. This is advantageous from the point of view of accurate standardization of the measurement method, and may permit the use of a cheaper indicating instrument.

(3) With a discharge time constant of 600 milliseconds, it is even more impractical to obtain a charging time constant of 1 millisecond than with a discharge time constant of 160 milliseconds.

(4) In the opinion of the author and of others who have used this type of radio-noise meter, the slower time constants are likely to result in indications which will correlate better with the annoyance factor of the radio noise as determined by listening experience.

(5) The longer time constants give a slower-moving, indicating-meter pointer, with consequent improved ease and comfort of reading.

Because of the adoption of the slower time constants for the rapid-response—slow-recovery circuit, there was no reason to follow the International Special Committee on Radio Interference in the specification of the indicating instrument, nor was there need to specify this instrument closely since its characteristics have thus purposely been made relatively unimportant in determining the response characteristics of the radio-noise meter. Therefore the specification of the indicating instrument has been broadly drawn to permit the use of standard, easily available commercial instruments.

Fig. 7—RCA Type 312A radio-noise meter showing rod antenna and carrying strap.

The use of a logarithmic scale is believed to be a very significant American advance over the International Special Committee on Radio Interference practice. The range of 100:1 is thought to be adequate to keep indications of rapidly varying noises "on scale," and provides a scale which can be read with greater accuracy than one of the same size with a range of 1000:1.

The International Special Committee on Radio Interference selectivity specification has been adopted for the Joint Coordination Committee specifications.

To permit field-intensity measurements, audible monitoring with headphones, or the connection of apparatus for the analysis of noise wave forms, it is specified that a switch be provided for converting the rapid-response—slow-recovery circuit to a conventional audio detector for supplying audio output terminals. As an example of the usefulness of such terminals, the overvoltage timer described by E. T. Dickey and the author[19] might be connected to them and used to determine the frequency of occurrence of peak noise voltages of various magnitudes.

A radio-noise meter, newly developed in accordance with these latest specifications of the Joint Coordination Committee, has recently been placed on the market. Figs. 6, 7, and 8 show three views of this instrument. It will tune to frequencies from 150 to 350 kilocycles and from 540 kilocycles to 18 megacycles in four bands, and will indicate from 10 microvolts to 100 millivolts in three direct-reading scales of 100:1 range each. Thus a wide overlapping of scales is provided to avoid the embarrassment of sometimes having to measure a rapidly varying noise on two different scales. Although standard batteries, readily procurable everywhere, are used, this instrument is strictly portable, with a weight, including batteries, of 32 pounds. A more complete description has been given by the author elsewhere.[20]

Radio-noise measurements in Canada have been made in accordance with the practices of both England and the United States. Because of close trade relationships with both England and the United States, Canadians are more disturbed by differences in these practices than are we in the United States. The Canadian Engineering Standards Association has been working toward the adoption of a Canadian standard for radio-noise meters which would follow as closely as possible both practices; and has tentatively specified an instrument in accordance with the Joint Coordination Committee 1940 report[18] but with the International Special Committee on Radio Interference or English time constants of 1 and 160 milliseconds. Listening tests were conducted in Toronto in May, 1940, by the Canadian Engineering Standards Association to determine, if possible, which of the two sets of time constants should be finally adopted. The results were inconclusive, however, because the time constants of the instruments were found not to be what had been intended.

Also in May, 1940, shortly after the Canadian tests, similar subjective correlation tests were conducted in New York by D. E. Foster of the RCA License Laboratory with the co-operation of the Joint Coordination Committee. Two radio-noise meters conforming to the Joint Coordination Committee specification, and one in accordance with the tentative Canadian specification were used in listening tests with three types of noise, produced by a commutator motor, an electric razor, and by a relay (clicks). The average variation in indication for the same subjectively determined quality of reception, for different types of noise, was found to be 2.9 and 7.7 decibels, respectively, for the Joint Coordination Committee instruments and for the Canadian instrument. The two Joint Coordination Committee instruments, which were of different manufacture, gave substantially equivalent performance.

These tests supply the best quantitative data on the extent to which the Joint Coordination Committee specification achieves its purpose of providing for radio-noise measurements which are reproducible with different instruments and which are indicative of the annoyance factor of different types of noise. A very satisfactory approach to this objective is indicated.

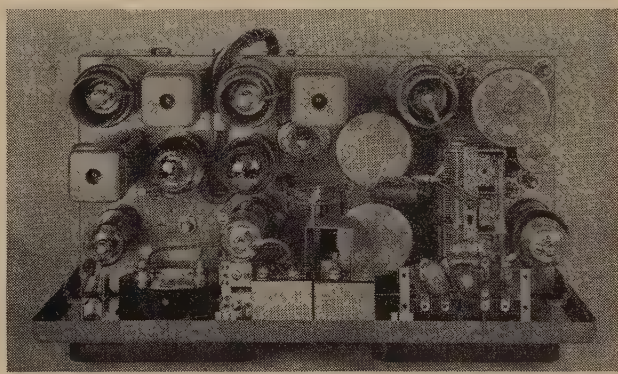


Fig. 8—RCA Type 312A radio-noise meter showing chassis withdrawn from its case.

In conclusion, the theoretical argument in favor of the use in noise measurements of an indicator of the rapid-response—slow-recovery type will be stated. The best published discussion of this matter known to the author is by A. H. Davis.[5] It has been shown by U. Steudel[21] that the response of the ear to single impulsive sounds is proportional to the peak value of the disturbance, and that the loudness of repeated impulses with a repetition rate up to 50 cycles per second is proportional to the repetition rate. By a proper choice of time constants, the rapid-response—slow-recovery type of indicator may be made to function in such a way as to approximate this action of the ear. This has been the intent in the latest International Special Committee on Radio Interference and Joint Coordination Committee radio-noise-meter specifications, although there is no great weight of experimental evidence as to how closely the desired result has been obtained.

Since such indicators of the rapid-response—slow-recovery type are definitely not intended to be peak

[19] C. M. Burrill, and E. T. Dickey, "The overvoltage timer and an example of its application to the measurement of radio interference," *R.M.A. Eng.*, vol. 3, pp. 16–21; November, 1938.

[20] Charles M. Burrill, "New equipment to measure intensity of radio noise," *Broadcast News*, no. 32, pp. 6, 7, 31, 34; March, 1940.

[21] U. Steudel, "Über Empfindung und Messung der Lautstärke (The sensation of loudness and its measurement)," *Hochfrequenz. und Elektroakustik*, vol. 41, pp. 116–128; April, 1933.

reading in all cases, it is suggested that they be called quasi-peak indicators. Doubtless experience eventually will show what are the best time constants for the quasi-peak indicator in a noise meter, but a thorough and well-planned program of research ought to be undertaken by someone, in order to determine these constants more directly and promptly. A. H. Davis[5] in his paper on the development of his acoustic-noise meter has set an excellent example in this respect which should be followed in the case of radio-noise meters.

The Design of the Universal Winding*

L. M. HERSHEY†, MEMBER, I.R.E.

Summary—*A simple gear-ratio formula for winding universal coils is derived. Auxiliary tables and curves to simplify its application for types of coils ordinarily used in radio work are also included.*

THE universal winding is quite widely used wherever a maximum inductance must be attained in a minimum volume. It is not generally superior in electrical characteristics to the single-layer solenoid, progressive universal, or bank windings. The factors influencing the choice of the type of coil are beyond the scope of this paper.

Before discussing the method of winding a universal coil it is desirable to describe the operation of the winding machine. Several varieties of machines are now in use; that shown in Fig. 1 is one of the simplest.

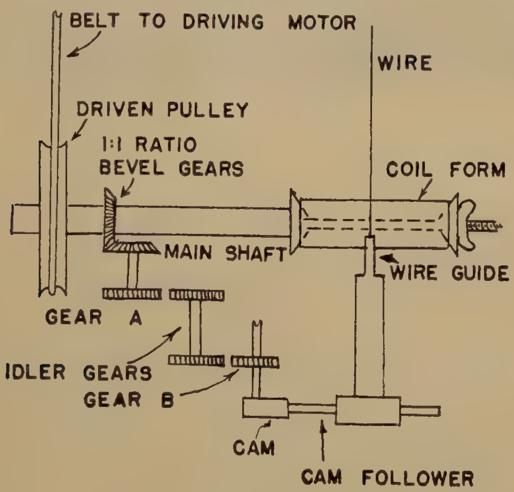


Fig. 1—Winding machine.

The main shaft is driven by a motor through a belt; the coil form is firmly attached to this shaft and rotates with it. A set of 1/1 ratio bevel gears drives gear A, which drives the idler gears. Generally, both idler gears have the same number of teeth, but sometimes another ratio such as 1/2 is used when a suitable gear ratio cannot be obtained using the 1/1 idler.

The idler gears drive gear B and the cam which is on the same shaft. The cam moves the wire guide back and forth axially across the periphery of the coil form (or coil) as the form turns. The cam is cut so that the wire guide moves at constant speed in a direction parallel to the axis of the coil form between cam reversals.

The wire is fed from a spool attached to a braking device which produces uniform wire tension, and is pulled off the spool and through the wire guide by the rotation of the coil form.

Development of the Gear-Ratio Formula

In the following discussion the coil form will be shown as if it had been slit lengthwise and rolled flat, and only that part of the form which is covered by the wire will be shown. So, for a simple example of a universal winding which is wound with a gear ratio (gear A/gear B) of 1/1, the pattern shown in Fig. 2 will result on the first turn. The total length of the pattern is π times the diameter of the form, and the width of the pattern is the throw of the cam c, neglecting the diameter of the wire used. Actually, of course, the total width covered by the winding will be $c+w$, where w is the diameter of the wire.

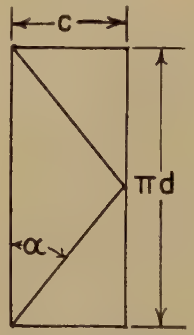


Fig. 2—Winding pattern.

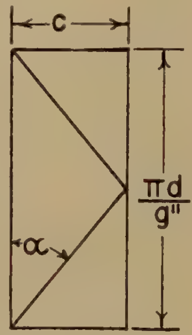


Fig. 3—Determination of g''.

If the gear ratio A/B is g'' in Fig. 3, one cam cycle will occur in the circumferential distance $\pi d/g''$, and

$$\tan \alpha = \frac{2cg''}{\pi d} \tag{1}$$

or

$$g'' = \frac{\pi d \tan \alpha}{2c} = \frac{1.57 d \tan \alpha}{c}. \tag{2}$$

When the cam reverses, the wire will slide if the winding angle α formed between the wire and the side of the coil is too large, and the coil will not wind to the proper width; the first few turns may be parallel all

* Decimal classification: R230×R382. Original manuscript received by the Institute, May 15, 1941.
† Hazeltine Service Corporation, New York, N. Y.

the way around. At any rate, it is probable a poor mechanical result will be obtained. On the other hand, if the winding angle is too small the coil will not build up with straight sides, and if the winding angle is extremely small, a random-wound coil will result. The tendency of the wire to slip on the coil form depends also on the diameter of the form, the coefficient of friction between the wire and the form, the amount of tension on the wire, and the stiffness of the wire itself.

The maximum value of tan α is a function of the form diameter for any given form material. Fig. 4 shows experimentally determined curves of this function for smooth (glazed bakelite) and fairly rough (cardboard) forms. The value of tan α can be read off the appropriate curve and inserted in (2) to calculate g''.

In a large number of cases a winding angle of 12 degrees has been found to be satisfactory. This value of α is particularly convenient to use since it results in the following special form of (2) which is simple and easy to remember:

$$g'' = \frac{d}{3c}. \tag{3}$$

The foregoing discussion has dealt with the choice of a gear ratio to produce the proper winding angle. The value of gear ratio also will determine the pattern the wire will describe during the winding cycle; the winding cycle is a whole number of cam cycles which starts and ends over nearly the same point on the coil form. The simpler fractions, resulting when small whole numbers are used in the gear ratio, produce simple winding patterns and usually a better mechanical result than that obtained when larger numbers are used. Therefore, it is desirable, when g'' as calculated from (2) or (3) is not a simple fraction, to choose the nearest simple fraction. This chosen value g' must be *smaller* than the calculated value g'', if the latter was computed from the experimentally determined *maximum* value of tan α.

In order to simplify the calculations involved in actually designing coils, values of g'' have been computed and the appropriate values of g' tabulated for most of the commonly used cams and form diameters. The values for smooth forms are given in Table I while those for rough forms appear in Table II.

TABLE I

Suggested Values of g' for Smooth Forms

Inside Diameter of Coil	Cam Throw					
	1/16 inch	3/32 inch	1/8 inch	3/16 inch	1/4 inch	3/8 inch
(inches)						
1/4	1/1	2/3	1/2	1/3	1/4	1/6
3/8	3/2	1/1	2/3	1/2	2/5	1/4
1/2	2/1	1/1	1/1	2/3	1/2	1/3
5/8	2/1	3/2	1/1	2/3	1/2	2/5
3/4	5/2	3/2	1/1	2/3	2/3	2/5
7/8	3/1	2/1	3/2	1/1	2/3	1/2
1	3/1	2/1	3/2	1/1	2/3	1/2
1-1/4	4/1	5/2	2/1	1/1	1/1	2/3

The change from the calculated value g'' to the chosen value of g' will result in a change in the winding angle at the diameter which was used in (2), and may cause a serious error in the equations which will be developed later, if some form of compensation is not added. Therefore, at this point a new and fictitious

TABLE II

Suggested Values of g' for Rough Forms

Inside Diameter of Coil	Cam Throw					
	1/16 inch	3/32 inch	1/8 inch	3/16 inch	1/4 inch	3/8 inch
(inches)						
1/4	2/1	1/1	1/1	2/3	1/2	1/3
3/8	3/1	2/1	3/2	1/1	2/3	1/2
1/2	3/1	5/2	3/2	1/1	2/3	1/2
5/8	4/1	3/1	2/1	3/2	1/1	2/3
3/4	5/1	3/1	5/2	3/2	1/1	2/3
7/8	5/1	3/1	5/2	3/2	1/1	2/3
1	6/1	4/1	3/1	2/1	3/2	1/1
1-1/4	7/1	5/1	3/1	5/2	3/2	1/1

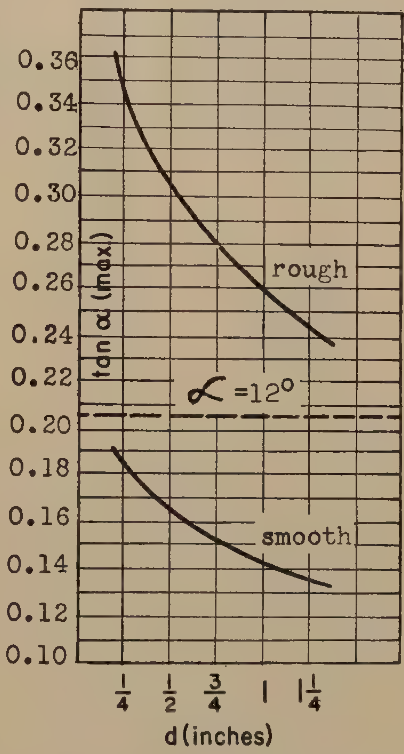


Fig. 4—Tan α versus form diameter.

diameter d' will be introduced. This is the diameter at which the cam and chosen gear ratio g' would produce the desired winding angle; it will be less than the actual inside diameter of the coil if the chosen value of g' is less than the calculated value. Equation (2) may be rewritten to determine this fictitious diameter

$$d' = \frac{2cg'}{\pi \tan \alpha}. \tag{4}$$

When g' is expressed as a fraction in its lowest terms, (using the smallest whole numbers for the numerator and denominator), some significant facts concerning the geometry of the winding cycle are apparent. If the numerator and denominator of g' are q' and s', respectively, q' is the number of cam cycles per winding cycle, and $(s'-1)$ is the number of lines apparent on the periphery of the coil; the pattern can be sketched by marking off $2q'$ spaces along the circumference of

the coil and s' spaces across the width, then drawing diagonal lines through the rectangles thus formed. It is necessary to use $2q'$ spaces on the circumference, since the cam crosses the coil in 1/2 cycle. This method was followed in developing the patterns of Fig. 5.

The patterns which have been developed up to this point would be repeated on the second and third suc-

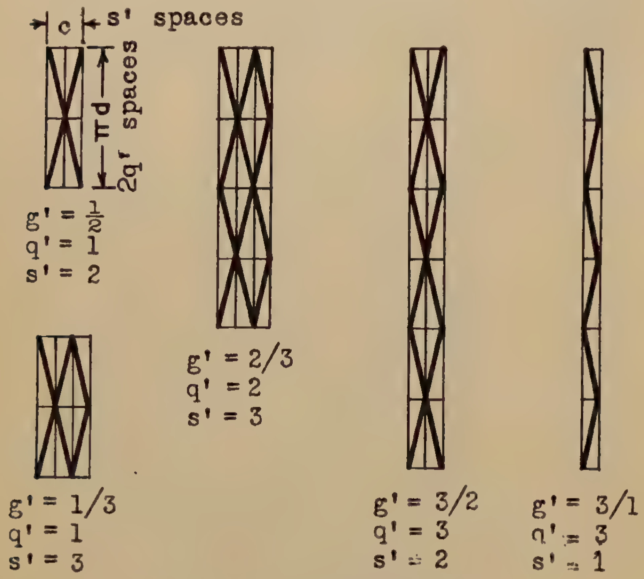


Fig. 5—Winding patterns.

cessive winding cycle, directly over the preceding winding-cycle pattern, with cam reversals occurring on the same radii and a decreasing winding angle as the coil grows in diameter. This does not produce the type of winding which is desired. If the first winding cycle is completed in just slightly more (a progressive winding) or slightly less (a retrogressive winding) than a whole number of turns of the coil, then each successive cam cycle will be slightly displaced from a similar cam cycle in the preceding winding cycle and the typical universal winding will be produced. Experience has indicated that the best mechanical results are obtained when the distance between centers of adjacent turns of wire is about 1.25 w, where w is the diameter of the wire. This displacement of successive winding cycles is obtained by using a gear ratio g which differs slightly from the simple ratio g'. The tightness of the winding is rather critically dependent on this difference, so it is necessary to derive an equation for g in terms of the various coil parameters.

Fig. 6 shows the part of a winding-cycle pattern produced by one cam cycle when the retrogressive type of winding is used, with the dotted line representing the pattern which would result if the gear ratio g' were used instead of g.

In the example shown, the winding starts at a, and the first cam cycle of the first winding cycle is completed at b. If g is the gear ratio used, the circumferential distance $a-b$ is $\pi d'/g$ or $1/g$ revolutions of the coil form; this differs from $\pi d'/g'$ by the circumferential distance x/q', which will be chosen so that the distance between adjacent wires is 1.25 w. The new winding angle β is slightly greater than α in the retrogressive winding or slightly less than α in the progressive winding. In Fig. 6, g is shown to be greater than g'. A similar diagram can be developed for $g<g'$.

The gear-ratio formula can now be written

$$\frac{\pi d'}{g} = \frac{\pi d'}{g'} \pm \frac{x}{q'} \tag{5}$$

where the use of the minus sign produces a retrogressive winding, and the plus sign produces a progressive winding. Equation (5) can be written in the more convenient form

$$\frac{1}{g} = \frac{1}{g'} \pm \frac{x}{\pi d' q'}\cdot \tag{6}$$

Fig. 7 shows the first cam cycle of a winding cycle $\overline{a-b}$ and the first cam cycle of the next winding cycle, $\overline{b'-e}$. It has been shown that there are q' cam cycles per winding cycle, so the total progression or retrogression of the cam cycle in a winding cycle $a-b'$ is q' times the progression or retrogression per cam cycle, or

$$\overline{a-b'} = q'\frac{x}{q'} = x. \tag{7}$$

It was stated that the distance between the centers of adjacent turns of wire is 1.25 w. This is the side $a-y$

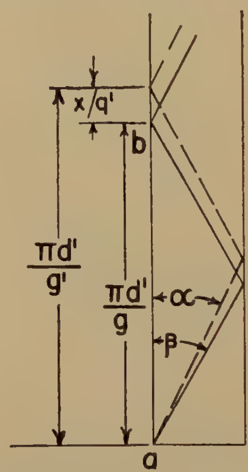


Fig. 6—One cam cycle of a retrogressive winding.

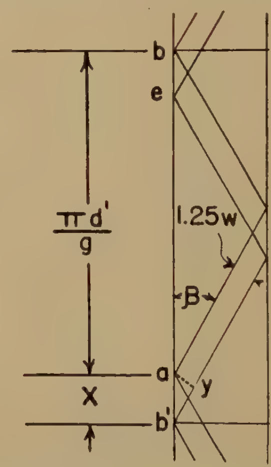


Fig. 7—Similar cam cycles of successive winding cycles.

of the right triangle $a-b'-y$, and the side $a-b'$ is equal to x.

$$x = \frac{1.25w}{\sin\beta} \tag{8}$$

and (6) can be rewritten, with this value of x,

$$\frac{1}{g} = \frac{1}{g'} \pm \frac{1.25w}{\pi d' q' \sin\beta}\cdot \tag{9}$$

The difference between the angles α and β is slight enough to be neglected in most types of radio coils, so

it may be assumed that they are equal and α substituted for β in (9).

$$\frac{1}{g} = \frac{1}{g'}\left(1 \pm \frac{1.25wg'}{\pi d'q' \sin \alpha}\right). \tag{10}$$

From (4)

$$d' = \frac{2g'}{\pi \tan \alpha},$$

so, substituting this value in (10),

$$\frac{1}{g} = \frac{1}{g'}\left[1 \pm \frac{1.25wg'(\pi \tan \alpha)}{2\pi cg'q' \sin \alpha}\right]. \tag{11}$$

For the small values of α involved, the error involved in calling the sine equal to the tangent is negligible; therefore (11) simplifies to

$$\frac{1}{g} = \frac{1}{g'}\left(1 \pm \frac{0.63w}{cq'}\right). \tag{12}$$

Ordinarily w/cq' is much smaller than 1, and the error is small if (12) is rewritten

$$g = g'\left(1 \mp \frac{0.63w}{cq'}\right). \tag{13}$$

The plus sign produces a retrogressive winding and the minus sign produces a progressive winding. The progressive winding has slightly closer spacing between adjacent turns of the same layer and the retrogressive winding slightly wider spacing than the 25 per cent of the wire diameter which was used to develop the formula.

Because of the interrelationships between the various constants, there are a number of other transformations which can be applied to (10). Some of these result in equations which are simpler algebraically than (13), but the latter is much better adapted to slide-rule solution. It is well to point out that the slide rule seems to be the only convenient means of picking the actual gears involved. The end of the *C* scale is set at the solution of (13) on the *D* scale; the various usable gear combinations can then be picked from the points of coincidence of the whole number divisions on the two scales. The gears normally in use have from about 25 to about 90 teeth. For very high or low values of g, it is usually necessary to use a 2/1 or 1/2 idler in the winding machine to utilize gears in the above-mentioned range of teeth.

Equation (13) appears to be accurate enough for most types of practical radio coils. It has been used with wire diameters from about 0.006 to 0.02 inch, on dowels from 1/4 to 1 1/4 inches in diameter, and with cam throws of from 1/16 to 3/8 inch. All of the gear ratios calculated using this equation have wound satisfactorily, provided that the cam throw was at least six times the wire diameter so that each layer of the winding provided a firm base for the next layer to build up on. If the cam throw is too small the turns tend to slide out of their proper position.

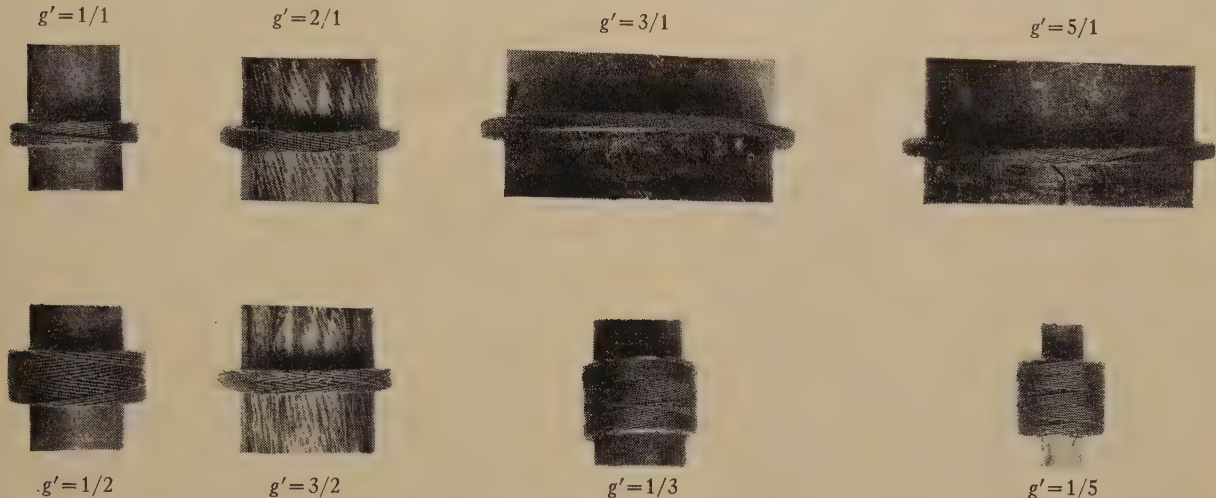


Fig. 8—Coil photographs.

If there is appreciable play in the mechanism of the winding machine, difficulty may be encountered in winding narrow coils with gear ratios having high values of g. The inaccuracy of each cam cycle is multiplied by g in each winding cycle and, therefore, the coil will build up properly only if the machine is in excellent mechanical condition.

There are several other factors which should be considered when starting the design of a winding. The details of the design of a coil for particular electrical characteristics are beyond the scope of this paper, but some general considerations will be dealt with briefly. Narrow coils usually have lower distributed capacitance than wide coils of the same inductance and therefore when low distributed capacitance is desired a small cam throw is used. A maximum electrical efficiency (Q) is generally obtained when the whole winding is about square in cross section; this is true whether the coil consists of one or more sections. Multisection coils are used when very low distributed capacitance is desired.

Fig. 8 is a collection of photographs of coils of various shapes and sizes, showing the appearance resulting from several commonly used values of g'.

Acknowledgment

The writer is indebted to Mr. N. P. Case and Professor Knox McIlwain for their interest and suggestions, which aided materially in the preparation of this paper.

Appendix

Calculation of the Winding Height

If the height of the winding above the form is to be known with a fair degree of accuracy it is necessary to calculate the number of turns per layer. Fig. 9 shows a part of one layer of a universal winding, with the lateral distance between centers of adjacent turns 1.25 times the wire diameter. The axial distance between similar points of adjacent turns is

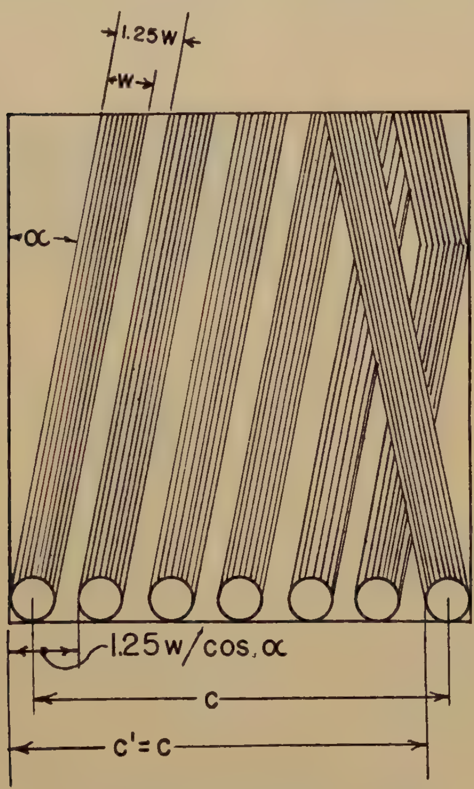


Fig. 9—One layer of a universal winding.

$$\frac{1.25w}{\cos\alpha} \approx 1.25w \tag{14}$$

for small values of α, and in the distance c there are thus

$$\frac{c}{1.25w} \text{ turns.}$$

By inspection of Fig. 9 it is obvious that the number of turns per layer is

$$\frac{c}{1.25w} + 1 = \frac{t}{m} \tag{15}$$

where t is the total number of turns on the coil and m is the number of layers.

The number of layers thus is

$$m = \frac{1.25wt}{c + 1.25w}. \tag{16}$$

The maximum height h of the winding above the form is

$$h = mw = \frac{1.25w^2t}{c + 1.25w}. \tag{17}$$

The coil would build up to a height mw if the wire were not flattened during winding. The amount of compression of the wire varies widely for different types of wire and insulation and no attempt will be made to evaluate it exactly. It results in a reduction of the winding height of roughly 5 to 10 per cent for most coils.

The Solution of Unsymmetrical-Sideband Problems with the Aid of the Zero-Frequency Carrier*

HAROLD A. WHEELER†, FELLOW, I.R.E.

Summary—*Unsymmetrical-sideband problems are met in frequency modulation and single-sideband transmission. There has developed the urgent need of a simplified procedure for their solution. The "vector envelope" of a modulated signal is reviewed, with special attention to the simple cases of amplitude, phase, and frequency modulation, and of a single sideband. From this is developed the concept of the "zero-frequency carrier" with combined amplitude and angle modulation. The solution is valid for any carrier frequency much greater than the total width of sidebands in the signal, regardless of whether the carrier is present. It yields directly the envelope of the signal, as detected by a rectifier. It is applied to the general case of steady and transient modulation components. The simplified procedure is outlined in simple terms after rigorous derivation. It involves merely stating the input modulating signal relative to the zero-frequency carrier, putting it through the low-pass analog of the band-pass filter, and deriving the output modulating signal directly.*

* Decimal classification: R148. Original manuscript received by the Institute, April 11, 1941.

† Hazeltine Service Corporation, Little Neck, L. I., N. Y.

I. Introduction

IN THE simple case of amplitude modulation, the sidebands are symmetrical about the carrier frequency. If the modulation envelope is to be detected by a rectifier, the carrier frequency must be greater than the highest frequency of modulation. If the rectified output is to have the carrier and sideband components removed by a low-pass filter, the lowest sideband frequency must be greater than the highest frequency of modulation; in other words, the carrier frequency must be greater than twice the highest frequency of modulation.

On the other hand, a zero-frequency carrier may be

conceived, with sidebands of positive and negative frequencies. If the sidebands are symmetrical, this represents merely the modulating signal superimposed on a direct component. Therefore this concept has physical reality in the case of amplitude modulation. It has been most helpful in visualizing and solving the problems of carrier modulation. Just as the carrier frequency and sidebands are conceived as shifted along the frequency scale for comparison between the modulating signal and the modulated carrier, so can a band-pass filter centered on the carrier frequency be transformed to an analogous low-pass filter centered on zero frequency.

In the general case, the carrier has a combination of amplitude and angle modulation, so the sidebands are not symmetrical. This is exemplified by single-sideband transmission and by phase or frequency modulation, the latter being particular types of angle modulation. It is then impossible to realize the corresponding modulation of a zero-frequency carrier. This concept is worth developing, even though it is not a physical one, because the mathematical simplification is even more needed in the general case.

The generalized modulation of the zero-frequency carrier leads to relations which are simple and not unexpected. Angle modulation introduces an imaginary quadrature component which is analogous to the quadrature carrier component that accompanies angle modulation. In the zero-frequency analog, the quadratic sum of the real and imaginary parts is analogous to the carrier envelope. This simplifies the rectifying process in mathematical treatments, because there are no carrier-frequency terms to be discarded.

Sideband dissymmetry may be present in the input signal, or may be introduced by band-pass filters not symmetrical about the carrier frequency. Examples of such filters are found in single-sideband selectors and in the slope filters of frequency-modulation detectors.[1–8]

There is to be described herein a very simple procedure for solving unsymmetrical-sideband problems, based on the concept of a zero-frequency carrier. The carrier frequency is assumed so high that the modulation envelope is clearly defined by the carrier peaks. The only unusual rule in this procedure is the necessity for retaining the distinction between positive-frequency and negative-frequency sideband components around the zero-frequency carrier. This is because they are not necessarily equal in magnitude and they are not treated alike by the filters. This rule is inherently associated with the unreality of angle modulation of a zero-frequency carrier, and of filter discrimination between positive and negative frequencies.

The zero-frequency carrier, modulated in amplitude, or angle, or both, is analogous to the "vector envelope" of the actual modulated carrier. To visualize the vector envelope, the modulated carrier is represented by a rotating vector. This rotating vector is viewed by an observer rotating steadily at the carrier frequency. There is seen merely a vector which is stationary in the absence of modulation. This vector expands and contracts in response to amplitude modulation. Its angle varies with phase or frequency modulation, but at a rate much less than the angular velocity of the rotating carrier vector. This slowly fluctuating vector is the vector envelope of the modulated carrier, so-called because it shows not only the amplitude modulation of the carrier but also its angle modulation.

II. The Vector Envelope of a Modulated Carrier

The full appreciation of the zero-frequency carrier requires a complete understanding of the vector representation of modulated waves, leading to the concept of the vector envelope. This background is here to be reviewed as a further introduction.[9–15]

The vectors here involved are two-dimensional vectors, so they may be located in a fixed plane. Each vector may be represented by a complex number. The real and imaginary parts of the complex number are the lengths of the projections of the vector on the real and imaginary axes, which are mutually perpendicular lines in the plane. The translational position of the vector in the plane is undetermined, so one end of the vector may be located at the origin (the intersection of the two axes) or several vectors may be drawn end to end for determining their sum.

It is customary to use vectors to represent alternating currents and voltages, especially their amplitude

[1] Recent solutions of single-sideband and frequency-modulation problems in picture transmission are examples of the need for simplified procedures, especially in the computation of the response to transient modulation. References 2, 3, and 4 yield results in simplest form but not by the simplest procedure.

[2] R. Urtel, "Observations regarding television transmission by a single sideband," *Telefunken-Hausmitteilungen*, vol. 20, pp. 80–83; July, 1939.

[3] H. E. Kallmann and R. E. Spencer, "Transient response of single-sideband systems," Proc. I.R.E., vol. 28, pp. 557–561; December, 1940.

[4] Charles P. Singer, "A mathematical appendix to transient response of single-sideband systems," Proc. I.R.E., vol. 28, pp. 561–563; December, 1940.

[5] S. Goldman, "Television detail and selective-sideband transmission," Proc. I.R.E., vol. 27, pp. 725–732; November, 1939.

[6] H. Nyquist and K. W. Pfleger, "Effect of the quadrature component in single sideband transmission," *Bell Sys. Tech. Jour.*, vol. 19, pp. 63–73; January, 1940.

[7] C. W. Carnahan, "F-M applied to a television system," *Electronics*, vol. 13, pp. 26, 30–32; February, 1940.

[8] H. A. Wheeler and J. C. Wilson, "The influence of filter shape-factor on single-sideband distortion," Proc. I.R.E., vol. 28, p. 253; May 1940. (Summary only.)

[9] V. D. Landon, "An analysis of frequency modulation," unpublished report, 1929.

[10] E. H. Armstrong, "Frequency modulation," Proc. I.R.E., vol. 24, pp. 689–740; May, 1936.

[11] E. A. Laport, "Characteristics of amplitude-modulated waves," *RCA Rev.*, vol. 1, pp. 26–38; April 1937. (Effect of unsymmetrical sidebands.)

[12] Hans Roder, "Noise in frequency modulation," *Electronics*, vol. 10, p. 22; May, 1937.

[13] P. P. Eckersley, "Asymmetric-side-band broadcasting," Proc. I.R.E., vol. 26, pp. 1041–1092; September, 1938.

[14] E. C. Metschl, "Knowledge and use of amplitude, phase, and frequency modulation, *Elektrotech. Zeit.*, vol. 60, pp. 1357–1361; November 30, 1939; vol. 60, pp. 1395–1401; December 7, 1939.

[15] W. L. Everitt, "Frequency modulation," *Elec. Eng.*, vol. 59, pp. 613–625; November, 1940.

and phase relations in a system. Since a current or voltage is a single-dimensional quantity, some qualification is necessary to enable its representation by a two-dimensional vector. The usual explanation is that the current or voltage is one projection of the vector. If the projection on the real axis is chosen, the current or voltage is given by the real part of the complex number corresponding to the vector.

Why should a simple one-dimensional quantity, such as current or voltage, be complicated by the use of a two-dimensional vector representation? The reason is that a rotating vector is a simpler concept than a sinusoidal wave. The latter is always visualized as a projection of the former. In mathematical analysis, the theory of complex numbers makes it easier to manipulate vectors than sinusoidal waves. In the simplest case of a wave of constant amplitude and frequency, the corresponding length and angular velocity of the vector are simpler to describe than the time variation of the wave.

The vector has an added advantage in problems involving modulation. It carries continuously an identification of both amplitude and angle, whereas a modulated wave does not carry uniquely the identification of either. A single point on a wave does not identify either the amplitude or the angle of a wave, unless the other is known or assumed, whereas a single representation of a vector identifies both of these properties. Therefore, it is possible to identify amplitude and phase modulation, separately and independently, only in the case of a vector.

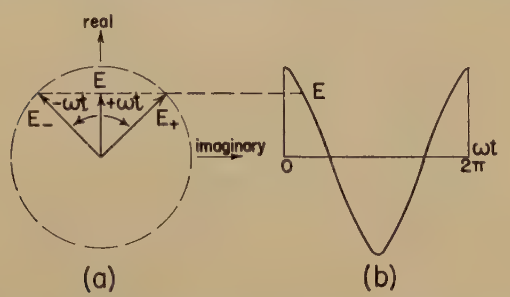


Fig. 1—The pair of oppositely rotating conjugate vectors (a) whose average or projection is a real cosine wave (b), one of these vectors being the usual representation of the cosine wave.

The vector or complex representation of an alternating wave may be justified from various points of view. The simplest is based on the fact that every equation in complex algebra is really two equations, one involving all real parts and the other involving all imaginary parts. The real parts being the essential quantities, it is sufficient that these quantities satisfy the equations. The imaginary parts are carried along merely to give the advantages of vector manipulation.[16]

Another representation of an alternating wave in terms of a vector involves also the conjugate of the vector. It happens that the real part or projection of a vector is the average of this vector and its conjugate vector.[17] The imaginary parts of two mutually conjugate vectors cancel out in their average, which is one half their sum; their real parts are equal to each other and therefore to the average of the two vectors. In complex notation, the conjugate is easily expressed. It is obtained merely by reversing the direction or sign of the angle.

In the case of a rotating vector, the angle includes ωt, the product of the radian frequency ω and the time t. The conjugate vector is the same except that it rotates in the opposite direction, and its angle includes $-\omega t$. In other words, the conjugate has a negative frequency.

It follows that an alternating wave is really the average of two conjugate vectors which are identical except for negative and positive frequencies. One is the image of the other, reflected in the real axis, which is consistent with their opposite directions of rotation. Any analysis which is to retain the real character of the wave, in spite of the use of vector notation, must recognize the two conjugate vectors of negative and positive frequency. It is not necessary to carry both vectors through all operations, because the result for one vector can usually be obtained by simple changes from the result for the other vector.

The conjugate vectors of both negative and positive frequency are here to be used in the representation of an alternating wave. Fig. 1(a) shows such a pair of vectors rotating in opposite directions at the same speed. They are represented by the complex expressions

$$E_- = A \exp - i\omega t = A(\cos \omega t - i \sin \omega t)$$
$$E_+ = A \exp i\omega t = A(\cos \omega t + i \sin \omega t). \tag{1}$$

The subscripts $(-)$ and $(+)$ denote respectively the negative and positive frequencies and directions of rotation, while A is the amplitude. The real alternating wave is the average of these two vectors:

$$E = \tfrac{1}{2}E_- + \tfrac{1}{2}E_+ = A \cos \omega t. \tag{2}$$

As far as the real wave is concerned, the frequency appears only under the cosine operator, so it is immaterial whether it is negative or positive. One or the other direction should be used consistently, the positive frequency being conventional. Fig. 1(b) shows the alternating wave corresponding to the real average value or the projection of the conjugate vectors.

A carrier with amplitude modulation is completely represented by three pairs of conjugate vectors, as shown in Fig. 2(a). One pair $E_{c\pm}$ is the carrier, whose radian frequency is denoted ω_0. The other two pairs $E_{r\pm}$ and $E_{s\pm}$ are the sidebands, whose respective frequencies are $\omega_0-\omega_m$ and $\omega_0+\omega_m$, ω_m being the modulation frequency. The carrier is assigned unit amplitude

[16] L. A. Hazeltine, "Electrical Engineering," Macmillan Company, New York, N.Y., 1924, pp. 178–181.

[17] E. A. Guillemin, "Communication Networks," John Wiley and Sons, New York, N.Y., 1931, vol. 1, pp. 70–75.

so the amplitude of each sideband is $m/2$, m being the modulation factor. The three negative-frequency vectors have the sum E_-, the three positive-frequency vectors E_+:

$$E_\pm = E_{c\pm} + E_{r\pm} + E_{s\pm}. \tag{3}$$

The carrier and two sidebands are represented by three pairs of conjugate vectors:

$$E_{c\pm} = \exp \pm i\omega_0 t$$

$$E_{r\pm} = \frac{m}{2} \exp \pm i(\omega_0 - \omega_m)t$$

$$E_{s\pm} = \frac{m}{2} \exp \pm i(\omega_0 + \omega_m)t. \tag{4}$$

The average of the two sum vectors is the real modulated-carrier wave,

$$E = \tfrac{1}{2}E_- + \tfrac{1}{2}E_+ = \cos \omega_0 t + \frac{m}{2} \cos (\omega_0 - \omega_m)t + \frac{m}{2} \cos (\omega_0 + \omega_m)t$$

$$= \cos \omega_0 t(1 + m \cos \omega_m t). \tag{5}$$

It is customary to represent such a modulated wave by one group of three vectors, stopping the rotation of the carrier wave and indicating only the rotation of the sideband vectors relative to the carrier, the remaining rotation being less rapid because the modulation frequency is less than the carrier frequency. This operation on the set of positive-frequency vectors is shown in Fig. 2(b). Their frequencies are all reduced by the amount of the carrier frequency. The carrier frequency becomes zero and the sideband frequencies $\pm\omega_m$. The added subscript (0) denotes the vectors of the group whose carrier frequency is shifted to zero:

$$E_{c0+} = 1$$

$$E_{r0+} = \frac{m}{2} \exp - i\omega_m t$$

$$E_{s0+} = \frac{m}{2} \exp i\omega_m t. \tag{6}$$

Their sum is

$$E_{0+} = 1 + m \cos \omega_m t. \tag{7}$$

It is noted that each of the three components in (6) is obtained by extracting the factor exp $i\omega_0 t$ from the corresponding component in (4). Since this factor is the carrier, the remaining factors in (6) are of the nature of modulation factors. Their sum (7) is the modulation envelope, plotted in Fig. 2(c), which pulsates at the modulation frequency ω_m. The sideband vectors are mutually conjugate, so the envelope is a real quantity.

The concept of the zero-frequency carrier is illustrated in this example, Fig. 2(b). By reducing the carrier frequency to zero, and the sideband frequencies by an equal amount, there is obtained a group of vectors whose sum is the envelope of the modulated-carrier wave. In other words, the envelope of a modulated zero-frequency carrier wave is identical with the wave itself.

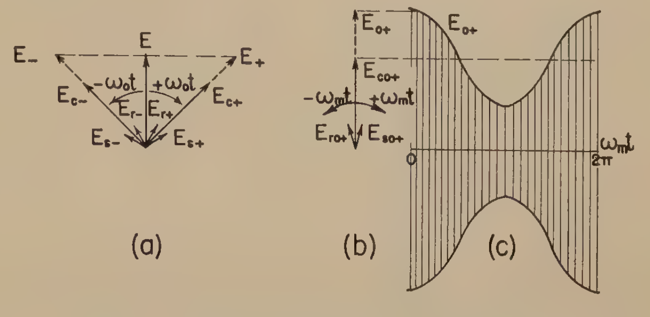


Fig. 2—The six rotating vectors (a) whose resultant is a real cosine wave with modulated amplitude, three of these vectors (b) being the usual representation of the carrier and two side frequencies which determine the modulation envelope (c).

In a carrier-wave communication system, the modulating signal in the transmitter is applied to the modulator as a modulated direct voltage, and that causes the modulator to yield the modulated carrier. In the receiver, the reverse process occurs in the detector, yielding again a modulated direct voltage. If a complete system is being studied, why not omit the processes of modulation and detection, performing all operations theoretically on a modulated zero-frequency carrier? It is the purpose of this treatment to deal with this question, particularly to give the conditions under which this is permissible and the rules which govern such a procedure. These rules have been familiar in the case of pure amplitude modulation, in which the sidebands are symmetrical, but have not been formulated for general modulation with unsymmetrical sidebands. The latter is the problem of phase or frequency modulation or of single-sideband operation.

As an introduction to unsymmetrical sidebands, the case of a carrier and a single sideband component is shown in Fig. 3. Each set of carrier and sideband vectors in Fig. 3(a) has a resultant $E_\pm$ which, in general,

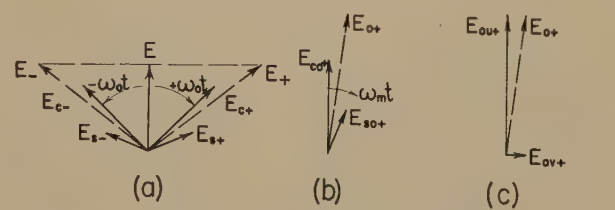


Fig. 3—The four rotating vectors whose resultant is a carrier and single sideband.
(a) The rotating vectors of carrier frequency and side frequency.
(b) The carrier vector stopped to show the relative phase displacement of side vector and resultant.
(c) The resultant analyzed into its real and imaginary components.

differs from the carrier vector $E_{c\pm}$ in both amplitude and phase. The carrier and single-sideband vectors are

$$E_{c\pm} = \exp \pm i\omega_0 t$$

$$E_{s\pm} = m \exp \pm i(\omega_0 + \omega_m)t. \tag{8}$$

Their sum may be expressed in terms of amplitude modulation a and phase modulation b:

$$E_{\pm} = E_{c\pm} + E_{s\pm} = (1 + a) \exp \pm i(\omega_0 t + b). \quad (9)$$

Expanding in trigonometric functions and identifying the coefficients of $\cos \omega_0 t$ and $\sin \omega_0 t$, the modulation terms are evaluated:[18,19]

$$1 + a = \sqrt{1 + m^2 + 2m \cos \omega_m t}$$

$$\tan b = \frac{m \sin \omega_m t}{1 + m \cos \omega_m t}. \quad (10)$$

These formulas express separately the amplitude and phase modulation inherent in the carrier and single sideband. The frequency of the composite signal is the time derivative of the angle of rotation:

$$\omega = \frac{d}{dt}(\omega_0 t + b) = \omega_0 + \frac{db}{dt}$$

$$= \omega_0 + m\omega_m \frac{m + \cos \omega_m t}{1 + m^2 + 2m \cos \omega_m t}. \quad (11)$$

The last term is the deviation from the carrier frequency.

In Fig. 3(b), the positive-frequency set of vectors is shown with a frequency reduction by the amount of the carrier frequency, so the carrier frequency becomes zero as in Fig. 2(b). In Fig. 3(b), however, the sum E_{0+} of the two resulting vectors E_{c0+} and E_{s0+} differs from the carrier vector in both angle and amplitude; it is not a real quantity but rather a complex quantity. Extracting the carrier factor in (8) and (9),

$$E_{c0+} = 1, \; E_{s0+} = m \exp ib, \; E_{0+} = (1 + a) \exp ib. \quad (12)$$

It appears that the vector E_{0+}, by its two-dimensional complex nature, contains both the amplitude and the angle modulation of the composite signal, but not the carrier frequency. This is what is termed the "vector envelope" of a modulated carrier which, in general, may have both amplitude and angle modulation. Since the frequency and phase of the carrier E_{c0+} are zero, the corresponding properties of the vector envelope E_{0+} are the frequency and phase deviation from the unmodulated carrier.

Unlike the real envelope of Fig. 2(b), the vector envelope of Fig. 3(b) does not have physical reality. It is impossible to modulate a direct voltage with respect to phase, because its only property is amplitude. Nevertheless, this concept is found useful. Its appreciation is not difficult if the stationary carrier is regarded as a zero-frequency vector, rather than a one-dimensional quantity such as direct current or voltage. The real and imaginary parts of the vector envelope may be subdivided as in Fig. 3(c):

$$E_{0+} = E_{0u+} + E_{0v+} \quad (13)$$

in which the subscripts u and v respectively denote the real or physical part and the imaginary or unphysical part. The envelope is real only if the sidebands occur in conjugate pairs, as in Fig. 2.

A passive detector, such as the usual diode peak detector, responds to the envelope amplitude without regard for frequency or phase. Therefore the amplitude or magnitude of the vector envelope determines the response of such a detector. Since the frequency of the unmodulated carrier does not affect the response, its omission is a logical simplification.

A comparison of the envelope amplitudes in the cases of double and single sidebands, Figs. 2 and 3, shows that the former but not the latter would yield an undistorted cosine wave as the output from a linear detector. In the single-sideband case, the envelope distortion decreases with the modulation factor m, the various kinds of modulation approaching the following forms for m much less than one:

$$a = m \cos \omega_m t, \quad b = m \sin \omega_m t, \quad \omega = m\omega_m \cos \omega_m t. \quad (14)$$

These are respectively the amplitude, phase, and frequency modulation. Just as Fig. 2 shows a pair of

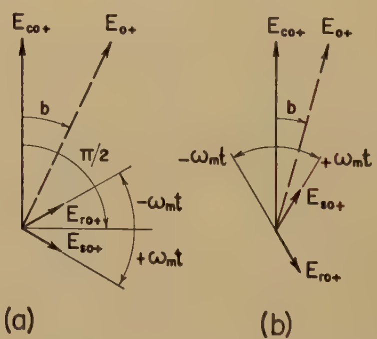


Fig. 4—Specially related carrier and sidebands for two cases of phase or frequency modulation.
(a) Maximum phase angle at initial time.
(b) Maximum frequency at initial time.

conjugate sidebands which represent pure amplitude modulation, there are pairs of nonconjugate sidebands which represent pure phase or frequency modulation. These are shown in Fig. 4, with a zero-frequency carrier.

In this further discussion of vector envelopes, the subscripts 0+ may be omitted because they are understood on all the vector quantities. The subscript 0 merely identifies the concept of zero-frequency carrier, while the + refers to the positive-frequency set of vectors.

Fig. 4(a) shows a pair of sidebands E_r and E_s which are not mutually conjugate with respect to the carrier E_c but are with respect to a vector perpendicular to the carrier. The sum of the sideband vectors is therefore perpendicular to the carrier, so its principal effect is to shift the angle of E, the sum of the carrier and

[18] M. G. Crosby, "Frequency modulation propagation characteristics," Proc. I.R.E., vol. 24, pp. 898–913; June, 1936, equation (3).
[19] M. G. Crosby, "Frequency modulation noise characteristics," Proc. I.R.E., vol. 25, pp. 472–514; April, 1937, equation (5).

sidebands. These three components of the vector envelope are

$$E_c = 1$$

$$E_r = \frac{m}{2} \exp i(\pi/2 - \omega_m t)$$

$$E_s = \frac{m}{2} \exp i(\pi/2 + \omega_m t). \tag{15}$$

Their sum is

$$E = 1 + im \sin (\pi/2 - \omega_m t) = 1 + im \cos \omega_m t$$
$$= (1 + a) \exp ib \tag{16}$$

in which

$$1 + a = \sqrt{1 + m^2 \cos^2 \omega_m t}$$
$$\tan b = m \cos \omega_m t. \tag{17}$$

The amplitude modulation a includes only the higher powers of m, the principal modulation being the angle b.

Fig. 4(b) shows another pair of sidebands which also are mutually conjugate about a vector perpendicular to the carrier. In this case, however, they are expressed as "skew-conjugate" about the carrier, which means conjugate except for a reversal of one relative to the other. These three components of the vector envelope are

$$E_c = 1$$

$$E_r = -\frac{m}{2} \exp - i\omega_m t$$

$$E_s = \frac{m}{2} \exp i\omega_m t. \tag{18}$$

Their sum is

$$E = 1 + im \sin \omega_m t = (1 + a) \exp ib \tag{19}$$

in which

$$1 + a = \sqrt{1 + m^2 \sin^2 \omega_m t}$$
$$\tan b = m \sin \omega_m t. \tag{20}$$

The frequency modulation is

$$\omega = \frac{db}{dt} = \frac{m\omega_m \cos \omega_m t}{1 + m^2 \sin^2 \omega_m t}. \tag{21}$$

A comparison of the cases of Figs. 2 to 4 is simplified by assuming the modulation factor m to be much less than one. The following tabulation gives a summary on this basis.

Figure	$a(am)$	$b(pm)$	$\omega(fm)$
2, am	$m \cos \omega_m t$	0	0
3, ssb	$m \cos \omega_m t$	$m \cos (\omega_m t - \pi/2)$	$m\omega_m \cos \omega_m t$
4(a), pm	0	$m \cos \omega_m t$	$m\omega_m \cos (\omega_m t + \pi/2)$
4(b), fm	0	$m \cos (\omega_m t - \pi/2)$	$m\omega_m \cos \omega_m t$

(22)

This table shows which cases have which kinds of modulation. It is well known that phase and frequency modulation go together, as different variations of "angle modulation." The table also shows that a single-sideband signal has both amplitude and angle modulation.

In the phase (pm) and frequency (fm) columns, all terms in the same column are the same except for the phase of the modulation wave. This is emphasized by expressing all terms in cosine form. The cosine form is the proper one to use for phase comparisons, because it is the average of conjugate exponential forms and is independent of the direction of rotation. This shows incidentally the right-angle phase lead of frequency relative to phase, which results from the definition of radian frequency as phase velocity. The added factor of the modulation-frequency coefficient ω_m in the frequency column also results from this definition.

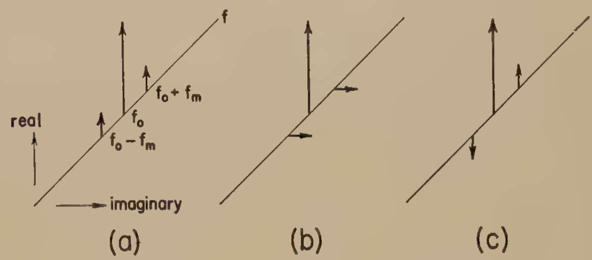


Fig. 5—Initial phase relations of carrier and sidebands.
(a) Amplitude modulation.
(b) Phase modulation.
(c) Frequency modulation.

The cases in the table have been selected to emphasize some peculiarities of certain sideband arrangements, especially those shown in Fig. 4. The case of Fig. 4(a) yields phase modulation with no phase shift of the modulation, while that of Fig. 4(b) similarly yields frequency modulation. Therefore these are the sets of sidebands which would be delivered from a phase or frequency modulator directly responsive to a modulating wave of standard phase, $\cos \omega_m t$. In this respect they correspond with the amplitude-modulation case of Fig. 2. The composite modulation in the single-sideband case of Fig. 3 appears to be amplitude and frequency modulation, as judged by the phase of the modulation.

Based on the three representative cases, Figs. 2, 4(a), and 4(b), there are shown in Fig. 5 the sideband relations uniquely corresponding to the three kinds of modulation. The single-sideband case is a combination of Figs. 5(a) and 5(c), the lower sideband canceling out.

The analysis of signals with any kind of modulation is simple with the concept of the zero-frequency carrier. Amplitude modulation of a direct current or voltage, as by a

microphone, is the well-known example. Angle modulation of a direct current or voltage is physically impossible, so it is merely a concept. It is not a difficult concept when the carrier is regarded as a zero-frequency vector. This vector is stationary in both amplitude and angle while unmodulated, but either or both of these properties can be subjected to modulation. If there were taken as a starting point the amplitude modulation of (7), the preceding process worked backward would then yield the corresponding set of sidebands. In the case of angle modulation, the process of deriving the sidebands produced by a given modulation is not so simple, but sometimes the identification of the sidebands is not required. If the amount of angle modulation is much less than one radian, a fair approximation of the sidebands can be obtained easily. The cases of Fig. 4 in the table (22) may be taken as a

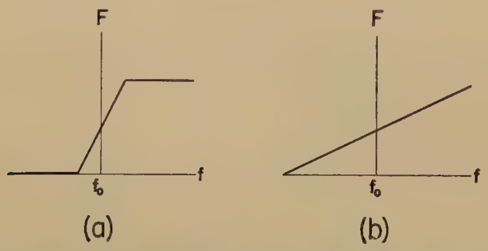


Fig. 6—Examples of unsymmetrical filters which disturb the relations between lower and upper sidebands.
(a) Single-sideband filter.
(b) Slope filter for a frequency detector.

starting point, in which case the corresponding sidebands are easily found to be those of formulas (15) and (18) and Figs. 4 and 5. The angle deviation is assumed to be so small that its sine is equal to the angle and its cosine is one.

Since the kind of modulation is determined by the relation between lower and upper sidebands, one kind of modulation may be changed to another by filters which modify this relation. In the case of a zero-frequency carrier, however, any physical filter cannot change one kind of modulation to another, because a physical filter has properties which prevent such a change. These properties are, with reference to zero frequency, the symmetric amplitude characteristics and the skew-symmetric phase characteristics. In other words, a physical filter cannot discriminate between negative and positive frequencies, just as a cosine wave cannot.

A physical filter can distinguish between lower and upper sidebands of a modulated carrier, providing the sidebands are relatively close to the carrier and differ therefrom by less than the carrier frequency. Fig. 6 shows two examples of filters for altering the amplitude relation between lower and upper sidebands about the carrier frequency f_0. Fig. 6(a) is a filter for dividing the sidebands, suppressing the lower sideband and doubling the amplitude of the upper sideband relative to that of the carrier. This may be termed a single-sideband filter, and is the type used in television systems. Fig. 6(b) is a slope filter for developing amplitude modulation in response to frequency modulation. It is the type used in frequency-modulation detectors.

The use of the rotating vector removes the confusion between negative and positive frequencies, so the carrier frequency can be reduced to zero, the lower sideband to negative frequencies and the upper sideband to lesser positive frequencies. Likewise, the unsymmetrical filters of Fig. 6 can be formulated around a carrier frequency f_0 of zero value. The fact that this is not a physical filter does not interfere with its use in theoretical derivations. If a real signal, with its conjugate sidebands, is put through such a filter, it comes out with an imaginary component from the sum of nonconjugate sidebands. This shows the effect of an unsymmetrical filter on the vector envelope.

The filters of Fig. 6 decrease one sideband in amplitude and increase the other. This tends to impart to a signal some of the properties of a single-sideband signal, especially the composite modulation of both amplitude and angle. This happens in the usual applications of these filters, namely, an amplitude-modulated signal through the filter of Fig. 6(a) or a frequency-modulated signal through that of Fig. 6(b). In the former case, no use is made of the frequency modulation. In the latter case, in a frequency detector, the resulting amplitude modulation is rectified to recover the signal carried by the frequency modulation.

The preceding examples are given to illustrate the identity of the vector envelope of a modulated carrier and the vector of a modulated zero-frequency carrier. The latter concept simplifies the formulation of any problem involving modulation of a carrier because the carrier frequency does not enter. The next section gives a generalized treatment of this concept and its limitations, its validity and its application, with special reference to transient signals.

III. The Generalized Problem of Carrier and Sidebands[20]

For practical purposes, a modulated-carrier signal has all of its sidebands within the band width bounded by 1/2 and 3/2 the nominal carrier frequency. This allows modulation frequencies not exceeding 1/2 the carrier frequency, so the modulation signal and the modulated-carrier signal occupy mutually exclusive frequency bands. This permits their separation by filters in modulators and detectors.

[20] At this point, it seems desirable to list the various subscripts, with their meanings:

$+, -$ denote the positive-frequency and negative-frequency components of a signal $E(t)$, or frequency ranges of a filter $F(f)$.
c, r, s denote carrier, lower-sideband, and upper-sideband components.
o, m denote carrier and modulation frequencies.
o also denotes vector envelope of carrier, or the low-pass analog of a band-pass filter.
u, v denote the real and imaginary parts of a complex quantity.
1, 2 denote the input and output signals.

Another way of stating the same condition is to say that the carrier frequency must be greater than twice the highest frequency of modulation, preferably much greater. In the case of transient modulation, a large sideband width is required, but it is always possible to make the carrier frequency still greater. Therefore this requirement is consistent with any practical conditions.

For theoretical purposes, it is convenient to deal with rotating vectors rather than sinusoidal oscillations. Any real sinusoidal wave is mathematically the sum of a pair of oppositely rotating vectors, one of positive frequency and the other of negative frequency. This convention is shown in Fig. 7(a). The modulated-carrier signal is separated into its positive-frequency and negative-frequency vector components. This applies to both carrier and sidebands. The expressed limitation of sideband width prevents overlapping of the sidebands of the positive-frequency carrier with those of the negative-frequency carrier. This simplifies the problem by preventing the interaction of sidebands from affecting the modulation envelope. Such an effect cannot occur in practice but does creep into theoretical derivations and tends to obscure the desired solution.

The modulated-carrier input signal of Fig. 7(a) is put through the band-pass filter of Fig. 7(b) to give the output signal of Fig. 7(c). One or both of the input and output signals may have unsymmetrical sidebands. The band-pass filter is a physical one, so its transmission at negative frequencies is equal in magnitude to that at positive frequencies. Its transmission is not necessarily symmetrical about the carrier frequency f_0. If not, it has a physical low-pass analog only in special cases.

For mathematical purposes, it is desirable that the positive-frequency modulated-carrier vector be not just one of two component vectors, but rather the vector conventionally used to represent the phase and amplitude of alternating voltages. Only half the amplitude would be had in one of the two component vectors. Instead, the signal is here regarded as the average of two oppositely rotating vectors, so each vector has the same amplitude as the signal and the positive-frequency vector has the same phase angle. In these terms, the input signal voltage, as a function of time, is

$$E_1(t) = \tfrac{1}{2}E_{1+} + \tfrac{1}{2}E_{1-} \tag{23}$$

in which E_{1+} and E_{1-} are the positive- and negative-frequency rotating vectors, represented by conjugate complex numbers.

The frequency spectrum of each voltage $E(t)$ is here represented by $F(f)$ with the same subscript. That of the input voltage is

$$F_1(f) = \tfrac{1}{2}F_{1+} + \tfrac{1}{2}F_{1-} = \int_{-\infty}^{\infty} E_1(t) \exp -i2\pi ft\, dt$$

$$F_{1\pm} = \int_{-\infty}^{\infty} E_{1\pm} \exp -i2\pi ft\, dt. \tag{24}$$

Strictly, the frequency spectrum is expressible only for transient signals or transient components of the signals, but it is also used loosely to represent any kind of signal. In the use of the Fourier integral for transient components of the signal, all steady components must be separated, then put through the filter individually, and finally recombined with transient components in the output.

The frequency spectrum, in order to include both amplitude and phase, has a real part for cosine coefficients and an imaginary part for sine coefficients. The magnitude of F is the resultant amplitude coefficient. For any real signal, the complex spectrum has only an even-real part symmetrical about zero frequency and an odd-imaginary part skew-symmetrical. This results from the physical identity of positive and negative frequencies.

Likewise it is desirable to express the filter transmission ratio in two parts, one for the positive-frequency pass band and the other for the negative. The positive-frequency band is the one conventionally used to represent the filter characteristics. In these terms, the filter transmission ratio, as a function of frequency, is

$$F(f) = F_+ + F_- \tag{25}$$

in which F_+ and F_- are the positive and negative-frequency band-pass characteristics. The filter ratio F has only even-real and odd-imaginary parts; that is, the real part is symmetrical about zero frequency and the imaginary part skew-symmetrical. The even part is representable by an even-power series and the odd part by an odd-power series. A physical filter can have only even-real and odd-imaginary parts, this limitation being related with its inability to distinguish between equal positive and negative frequencies. The positive- and negative-frequency parts F_+ and F_- have equal real parts and opposite imaginary parts.

Each part of the filter ratio, F_+ or F_-, may or may not be symmetrical about its carrier frequency $\pm f_0$. It does not have a physical low-pass analog unless it complies with the same conditions about the carrier frequency, as any physical filter does about zero frequency.

The output signal is derived from its frequency spectrum both expressed the same as the input except for a change of subscript:

$$F_2(f) = \tfrac{1}{2}F_{2+} + \tfrac{1}{2}F_{2-} = F_1F; \qquad F_{2\pm} = F_{1\pm}F_\pm \tag{26}$$

$$E_2(t) = \tfrac{1}{2}E_{2+} + \tfrac{1}{2}E_{2-} = \int_{-\infty}^{\infty} F_2(f) \exp i2\pi ft\, df;$$

$$E_{2+} = \int_0^{\infty} F_{2+} \exp i2\pi ft\, df;$$

$$E_{2-} = \int_{-\infty}^{0} F_{2-} \exp i2\pi ft\, df. \tag{27}$$

It is here noted that since F_{2+} and F_{2-} are respectively confined to the positive- and negative-frequency

regions, the integration for E_{2+} and E_{2-} need cover only these regions.

In a simple case like pure amplitude modulation, the carrier frequency is uniquely determined in the signal. Even if the carrier is absent, its frequency is the one about which the sidebands are symmetrical. In unsymmetrical frequency modulation, the sidebands may be unsymmetrical in such a way that the carrier frequency loses its identity during modulation, but is identified with its unmodulated value. In single-sideband transmission with carrier suppression, the carrier frequency loses its identity in the signal but must be communicated to the receiver or otherwise ascertained by the receiver operator. In the present treatment, the assignment of a carrier frequency f_0 is arbitrary as far as the mathematical procedure is concerned. Therefore it is assigned in accordance with the physical conditions in any given problem. These usually give a logical basis for the assignment of a carrier frequency, even if one is not present in the modulated signal.

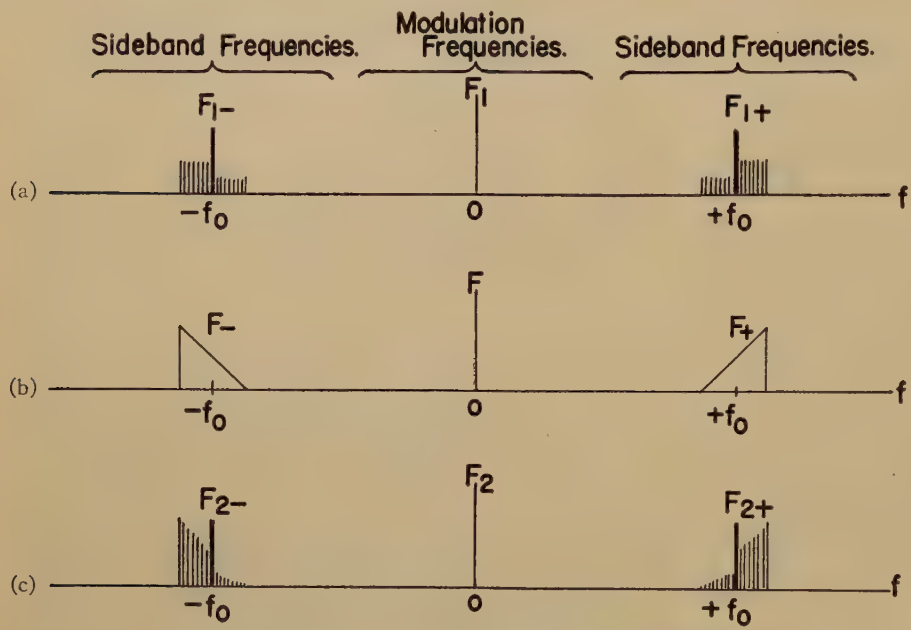


Fig. 7—The generalized problem of carrier and sidebands.
(a) The input-signal spectrum.
(b) The filter.
(c) The output-signal spectrum.

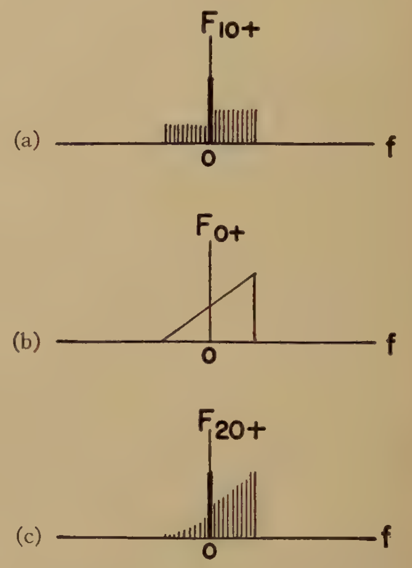


Fig. 8—The concept of zero-frequency carrier.
(a) The input-signal spectrum.
(b) The filter.
(c) The output-signal spectrum.

IV. The Concept of the Zero-Frequency Carrier

The frequency spectrum of a real modulated-carrier signal, or the band-pass character of a physical filter, is completely determined by its positive-frequency term, F_{1+}, F_+, or F_{2+} in (24), (25), or (26). Therefore a complete solution of the problem can be obtained without reference to the negative-frequency terms.

The positive-frequency terms are regarded as shifted along the frequency axis until the carrier frequency is zero, as shown in the transition from Fig. 7 to Fig. 8. The procedure to be described and verified permits the performance of all operations with the zero-frequency carrier. The actual carrier frequency f_0 does not even appear in the procedure or the final solution, but only in the proof of its validity.

The theoretical proof of this procedure is based on one of the well-known theorems of the Fourier integral.[21] Equation (23) may be restated in the form of a modulated-carrier signal:

$$E_1(t) = \tfrac{1}{2}E_{10+} \exp + p_0 t + \tfrac{1}{2}E_{10-} \exp - p_0 t \qquad (28)$$

in which E_{10+} and E_{10-} are the modulating voltages, in vector form, applied to the positive- and negative-frequency carrier vectors. In the present treatment, the various forms of the frequency variable are used interchangeably:

$$p = i\omega = i2\pi f; \qquad \omega = 2\pi f \qquad (29)$$

as in the tables of Campbell and Foster. The mentioned theorem yields the corresponding restatement of the frequency spectrum of (24):

$$F_1(f) = \tfrac{1}{2}F_{10+}(f - f_0) + \tfrac{1}{2}F_{10-}(f + f_0);$$

$$F_{10\pm}(f) = \int_{-\infty}^{\infty} E_{10\pm}(t) \exp - i2\pi ft\, dt \qquad (30)$$

in which F_{10+} and F_{10-} are the frequency spectrums of the modulating voltages as functions of $f \pm f_0$. Each of these spectra occupies a region around zero frequency, within $\pm 1/2$ the carrier frequency. Referring to the positive-frequency carrier terms,

[21] G. A. Campbell and R. M. Foster, "Fourier integrals for practical applications," Bell Telephone System Monograph B-584, September, 1931. Abridgment, *Bell Sys. Tech. Jour.*, vol. 7, pp. 639–707; October, 1928. (See Table I, No. 206.)

$$E_{1+}(t) = E_{10+} \exp p_0 t \tag{31}$$

$$F_{1+}(f) = F_{10+}(f - f_0). \tag{32}$$

The output signal is expressed in like terms with changed subcripts.

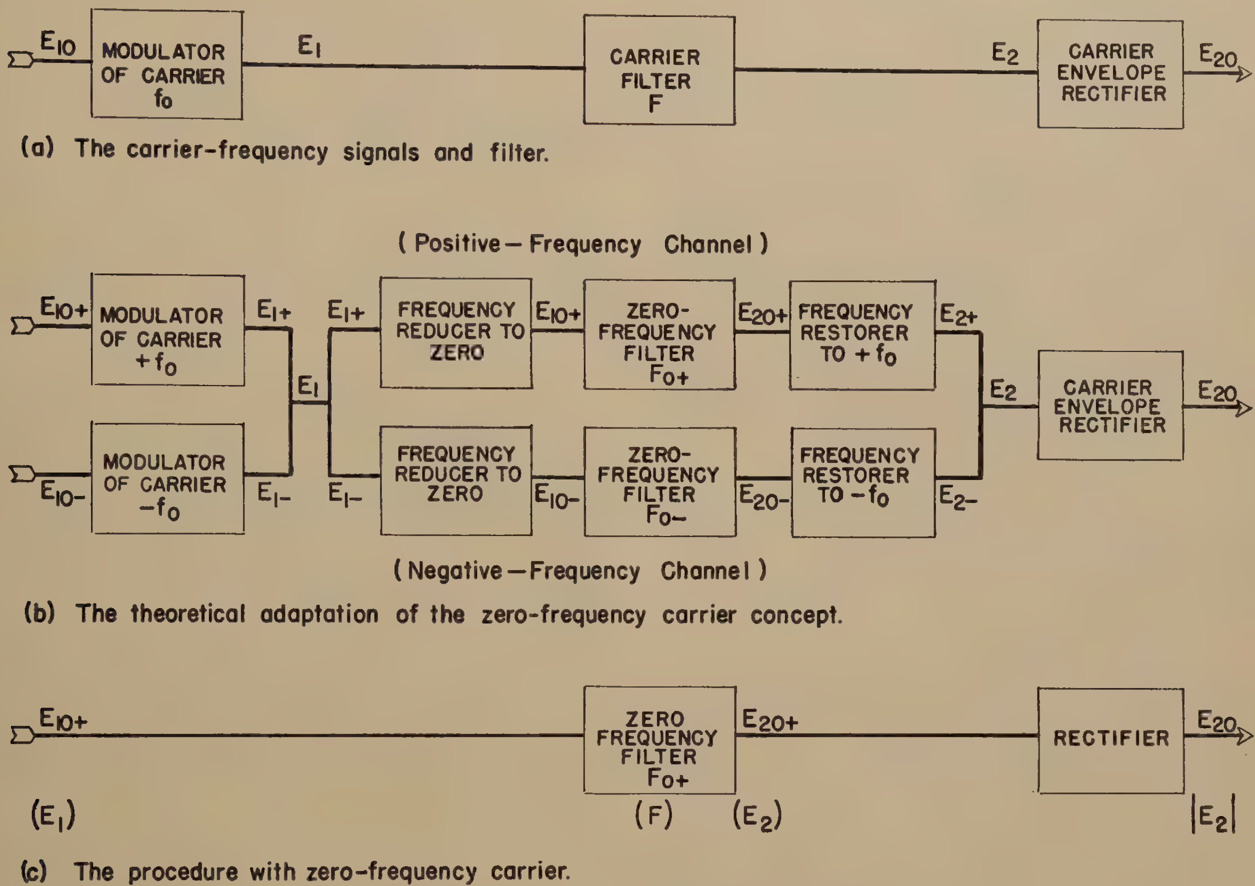


Fig. 9—The processes involved in the problem and its solution.
(a) The carrier-frequency signals and filter.
(b) The theoretical adaptation of the zero-frequency carrier concept.
(c) The procedure with zero-frequency carrier.

The character of the filter in general and in the positive-frequency pass band is expressed in the same manner:

$$F(f) = F_{0+}(f - f_0) + F_{0-}(f + f_0) \tag{33}$$

$$F_{+}(f) = F_{0+}(f - f_0) \tag{34}$$

in which F_{0+} and F_{0-} are the low-pass analogs of the band-pass filter.

Expressions (28) to (34) are the same in form as might be used in simple amplitude modulation. In the present more general treatment, however, special attention is required because the modulating voltages such as E_{10+} need not be real, and the frequency spectra and filter characters such as F_{10+} and F_{0+} need not be physical. This is the generalization involved in the treatment of unsymmetrical sidebands. There is a departure from reality in the zero-frequency carrier analog.

A few simple rules enable the formulation and solution to be made of any problems of symmetrical or unsymmetrical sidebands or band-pass filters in terms of the zero-frequency carrier without reference to the carrier frequency. The zero-frequency carrier has real and imaginary parts which determine both its amplitude and its phase. These are analogous to the amplitude and phase of the actual carrier. The envelope is determined only by the amplitude, the quadratic sum of the real and imaginary parts. With respect to the steady terms, special care is taken to retain the identity of positive and negative frequencies about the zero-frequency carrier, because they are not mathematically distinct. For example, the relation $\sin -\omega t = -\sin \omega t$ must not be used.

The rules regarding the zero-frequency carrier concept are here to be derived. Their use, however, is simple and does not require familiarity with their rigorous derivation.

V. The Mathematical Derivation

The problem is formulated in Fig. 9(a). A modulating signal E_{10} is applied to a carrier of frequency f_0 to give the modulated-carrier input voltage E_1. This is put through the band-pass carrier filter F to give the output voltage E_2. A rectifier is used to determine the envelope E_{20} of the output voltage. The generalized

problem is complicated by the fact that mutually independent amplitude and phase modulation cannot be produced by a real modulating signal E_{10}.

The mathematical steps in the proof are outlined in Fig. 9(b), while the derived procedure is given in Fig. 9(c). The assumptions and conditions are outlined in the preceding sections.

The generalized input signal E_1 is real. It is regarded as the product of a complex modulator capable of combined amplitude and phase modulation, determined by the respective real and imaginary components of the modulating signals E_{10+} and E_{10-}. The former of these signals is applied to the positive-frequency carrier vector and the latter to the negative, as indicated in (28) and Fig. 9(b). The average of these two signals determines the amplitude modulation and their differential average[22] the phase modulation.

It is easy to show that E_{10+} and E_{10-} are conjugate because the input signal E_1 is real. The real and imaginary parts are denoted respectively by the subscripts u and v:

$$E_{10+} = E_{10u+} + E_{10v+}; \quad E_{10-} = E_{10u-} + E_{10v-}. \tag{35}$$

Substituting in (28) and expanding the exponential terms representing the rotating carrier vectors,

$$E_1 = \tfrac{1}{2}(E_{10u+} + E_{10v+})(\cos \omega_0 t + i \sin \omega_0 t) + \tfrac{1}{2}(E_{10u-} + E_{10v-})(\cos \omega_0 t - i \sin \omega_0 t). \tag{36}$$

Equating to zero the imaginary coefficients of $\cos \omega_0 t$ and the real coefficients of $i \sin \omega_0 t$,

$$\begin{aligned} E_{10u+} - E_{10u-} &= 0; & E_{10u+} &= E_{10u-} \\ E_{10v+} + E_{10v-} &= 0; & E_{10v+} &= -E_{10v-}. \end{aligned} \tag{37}$$

Since their real parts are equal and their imaginary parts are opposite, E_{10+} and E_{10-} are conjugate. Therefore a knowledge of E_{10+} is sufficient to determine any real signal E_1. This signal has the form

$$E_1 = E_{10u+} \cos \omega_0 t + iE_{10v+} \sin \omega_0 t \tag{38}$$

and its envelope is

$$E_{10} = \sqrt{|E_{10u+}|^2 + |E_{10v+}|^2} = |E_{10+}|. \tag{39}$$

This is a proof that the complex modulating voltage E_{10+} is a representation of combined amplitude and phase modulation.

Passing the input signal E_1 through the filter multiplies its frequency spectrum by the filter factor to yield the frequency spectrum of the output signal E_2:

$$F_2 = F_1 F = (F_{1+} + F_{1-})(F_+ + F_-). \tag{40}$$

But, since these positive-frequency terms are zero in the negative-frequency region and vice versa, the product is simply

$$F_2 = F_{1+}F_+ + F_{1-}F_- = F_{2+} + F_{2-} \tag{41}$$

[22] The "differential average" of two quantities is half their difference, just as their average is half the sum. If the two quantities are conjugate complex numbers, their average is the real part of either and their differential average is the imaginary part of the one taken as positive.

and the parts of the product are

$$F_{2+} = F_{1+}F_+; \quad F_{2-} = F_{1-}F_-. \tag{42}$$

Expressing the positive-frequency spectrum of the output signal in the manner of (32),

$$F_{2+}(f) = F_{20+}(f - f_0). \tag{43}$$

From (32), (34), and (42),

$$F_{2+}(f) = F_{10+}(f - f_0)F_{0+}(f - f_0). \tag{44}$$

Therefore, changing the variable from f to $f + f_0$, and equating (43) and (44),

$$F_{20+}(f) = F_{10+}(f)F_{0+}(f). \tag{45}$$

This is the frequency spectrum of the modulating voltage E_{20+} of the positive-frequency vector representing the output signal:

$$E_{2+} = E_{20+} \exp p_0 t. \tag{46}$$

This process is essentially that outlined in the upper channel of Fig. 9(b). The input signal vector E_{1+} is shifted along the frequency scale to zero carrier frequency, and likewise the filter factor F_+. These are combined to give the modulating voltage E_{20+}, after which the carrier frequency is restored from zero to f_0, yielding the output signal vector E_{2+}. In the lower channel, a similar procedure is conceived. It does not have to be followed because the output E_{2-} is the conjugate of E_{2+}, since their average E_2 is real.

The envelope amplitude E_{20} of the output voltage E_2 is the magnitude of E_{20+}, as shown for E_1 in (38) and (39) above:

$$E_2 = E_{20u+} \cos \omega_0 t + iE_{20v+} \sin \omega_0 t \tag{47}$$

$$E_{20} = \sqrt{|E_{20u+}|^2 + |E_{20v+}|^2} = |E_{20+}|. \tag{48}$$

This is the output of the envelope rectifier.

The essential procedure for computations is very simple, as outlined in Fig. 9(c). The input modulating signal is put through the low-pass analog of the filter to yield the output modulating signal. The magnitude of this signal is the envelope obtained by rectifying the output. There remains only to summarize this procedure for practical use.

VI. The Procedure with the Zero-Frequency Carrier

In the solution of unsymmetrical-sideband problems with the zero-frequency carrier, there is no reference to the actual carrier frequency. The operations are performed on the modulating vectors. The filter is present in the form of its low-pass analog. In general, the modulating vectors are unreal and the low-pass filter unphysical, but the procedure is straightforward.

To reduce the number of subscripts, the 0 denoting "about the zero-frequency carrier" and the + denoting the positive-frequency channel in Fig. 9(b) are both omitted in all expressions to follow. The simplified symbols are noted under the diagram of Fig. 9(c).

The first step is the expression of the input modulating vector E_1. Its amplitude is the carrier amplitude. Its phase angle is that of the signal relative to the phase angle of the reference carrier. It has real and imaginary components as in (35).

The first step involves setting up the problem relative to a chosen carrier frequency. This is not a question of validity of the solution, but rather one of convenience. The proper expression relative to the nominal carrier frequency of zero is that which yields greatest simplification of the solution. If the carrier is identified in the signal, the problem is set up around this carrier frequency. In a balanced frequency detector, the nominal carrier frequency is most logically the frequency of balance. This step yields the vector E_1 as a complex number denoting the amplitude and phase modulation.

The second step is to separate the steady from the transient components, which may be denoted by subscripts s and t. These subscripts are here omitted and the steady terms are just mentioned in passing. They are separately passed through the filter F and are held until needed for recombination.

In all steps affecting steady components, it is essential to maintain the separate identity of equal positive and negative frequencies, because they are not alike in unsymmetrical sidebands and filters. Terms of equal positive and negative frequencies must not be combined in any manner which might lose their identity, until they emerge from the last unphysical filter.

The third step is to express the frequency spectrum of the transient signal E_1. This is done by means of the Fourier integral

$$F_1(f) = \int_{-\infty}^{\infty} E_1(t) \exp - i2\pi ft\, dt. \tag{49}$$

Many pairs of coefficients $F(f)$ and $E(t)$ are found in tables,[23] but some cases are better integrated directly.

The fourth step is to multiply the input frequency spectrum by the filter factor to obtain the output frequency spectrum

$$F_2 = F_1F. \tag{50}$$

The filter factor F is expressed relative to zero frequency instead of the carrier frequency.

The fifth step is to compute the output signal transient from the output frequency spectrum by means of the other form of the Fourier integral

$$E_2(t) = \int_{-\infty}^{\infty} F_2(f) \exp i2\pi ft\, df. \tag{51}$$

This may be obtained from the tables or integrated as needed.

The sixth step is to recombine the steady output terms with the transient terms obtained by the Fourier integral. This gives the complete modulating vector of the output signal. Therefore it is no longer necessary to avoid combination of positive- and negative-frequency terms.

If the modulation envelope is desired it is computed as the magnitude of E_2. The phase angle of the output modulation, which is the phase angle of E_2, may also be used in some problems.

Since many terms may be involved in this solution, it is sometimes useful to break it into parts which have physical significance, to facilitate manipulation and checking. Some useful divisions are here to be outlined.

The most useful line of division is between real and imaginary signal components, and between physical and unphysical terms of the frequency spectrum. Imaginary signal components correspond to unphysical terms of the frequency spectrum or filter factor, because real signals and physical filters can yield only real signals. A physical filter or frequency spectrum has only even-real and odd-imaginary terms. Therefore odd-real and even-imaginary terms are unphysical. The subscripts u and v are used respectively to denote the real and imaginary components of a signal or the physical and unphysical parts of a frequency spectrum or a filter factor.

In the first step, as in (35), the input voltage is divided into real and imaginary parts:

$$E_1 = E_{1u} + E_{1v}. \tag{52}$$

These are carried separately in the steady and transient components.

In the third step, the corresponding two parts of the input frequency spectrum may be obtained by separate integration in the same formula (49). The resulting physical and unphysical parts are

$$F_1 = F_{1u} + F_{1v}. \tag{53}$$

In the fourth step, the filter factor is separated into its physical and unphysical parts:

$$F = F_u + F_v. \tag{54}$$

The physical part (subscript u) contains all even-real and odd-imaginary terms, while the unphysical part (subscript v) contains all odd-real and even-imaginary terms. It is noted that even terms are symmetrical while odd terms are skew-symmetrical about zero frequency. The resulting output frequency spectrum is then

$$F_2 = (F_{1u} + F_{1v})(F_u + F_v). \tag{55}$$

It is easily shown that its physical and unphysical parts are respectively

$$F_{2u} = F_{1u}F_u + F_{1v}F_v; \quad F_{2v} = F_{1u}F_v + F_{1v}F_u. \tag{56}$$

In the fifth step, the physical and unphysical parts can be integrated separately to give the real part E_{2u} and imaginary part E_{2v}. Their quadratic sum is the output envelope.

[23] Refer to Campbell and Foster[21]. In these tables, $G(g)$ corresponds to the time function $E(t)$ used herein.

There are many useful rules in the handling of the frequency spectrum in relation to the signal. For example, the factor p applied to the frequency spectrum corresponds to a time differentiation of the signal, and $1/p$ to integration.[24] These rules are best learned by experience and by reference to the tables of paired coefficients.

VII. Conclusion

The concept of the zero-frequency carrier is shown to be useful in unsymmetrical-sideband problems, just as it has been in the simple case of amplitude modulation with symmetrical sidebands. This concept, with suitable adaptation, is adequate for the solution of problems with combined amplitude and phase modulation. The solution is independent of the actual carrier frequency and yields directly the modulation envelope of the output signal.

The full appreciation of this method requires some experience with practical examples. These have not been included here because it is planned to have them appear soon in special treatments of single-sideband and frequency-modulation problems.[25]

[24] Refer to Campbell and Foster[21], Table I, Nos. 208 and 210.

[25] H. A. Wheeler, "Common-channel interference from two frequency-modulated signals." Presented, Rochester Fall Meeting, November 12, 1940.

The Approximate Representation of the Distant Field of Linear Radiators*

RONOLD KING†, ASSOCIATE, I.R.E.

Summary—*After a brief résumé of the general theory of the distant field of arrays of linear radiators in terms of the vector potential (including the definition of a radiation function, radiation resistance, and directivity), approximate representations in terms of a sine distribution of current and in terms of an "equivalent" uniform distribution with an "effective length" are considered. It is shown that the conventional definition of an "effective height or length" $2h_e$ for a symmetrical, center-driven antenna of length $2h$, viz., $2Ih_e=\int_{-h}^{h}I_z dz$, does not give the best approximation. This is obtained by expanding the field characteristic in a Fourier series and defining a new "effective length" to be the coefficient of the leading term. It is shown that a reasonably good approximate representation in terms of an "effective length" is possible for antenna half lengths h which are less than λ/π, but not for longer ones. The case of two crossed antennas is discussed as an illustration of the simplification obtained using the approximate formulas.*

The General Theory of the Distant Field

The electromagnetic field of an antenna array consisting of one or more linear radiators of small radius a is conveniently calculated from the vector potential. At points in empty space this is given in volt-seconds per meter by

$$A = \frac{\Pi}{4\pi}\int_s I_s' \frac{e^{-j\beta R}}{R} ds'. \tag{1}$$

Here I_s' is the complex-current amplitude in amperes flowing in the element ds' (primed co-ordinates); R is the distance in meters from the point (unprimed co-ordinates) at which A is calculated to the element ds'; $\beta(=\omega\sqrt{\Pi\Delta})$ is the propagation constant in radians per meter; Π is the magnetic constant of space given numerically by $\Pi=4\pi\cdot 10^{-7}$ henry per meter; Δ is the electric constant of space given numerically by $\Delta=10^{-9}/36\pi$ farad per meter; the velocity c is defined by $c=1/\sqrt{\Pi\Delta}=3\cdot 10^8$ meters per second. The electric and magnetic fields may be calculated in general from (1) using the relations

$$E = (-j/\omega\Pi\Delta)(\text{grad div } A + \beta^2 A) \tag{2}$$

$$R = \text{curl } A \tag{3}$$

E is measured in volts per meter; B in volt-seconds per meter squared.

Expressions (1), (2), and (3) become simpler if the conditions defining the distant or radiation zone are imposed. These are

$$\beta R_0 \gg 1 \tag{4}$$

$$R_0 \gg \text{maximum dimension of array.} \tag{5}$$

R_0 is the distance from the point of calculation to a convenient reference origin 0 at the center of the array. Subject to (4) and (5) one has,

$$A^r = \frac{\Pi}{4\pi}\frac{e^{-j\beta R_0}}{R_0}\int_s I_s' e^{j\beta(R_0,s')}ds' \tag{6}$$

$$E^r = -j\omega\{\boldsymbol{\theta}_1 A_\theta{}^r + \boldsymbol{\Phi}_1 A_\Phi{}^r\} \tag{7}$$

$$B^r = -j\beta\{\boldsymbol{\Phi}_1 A_\theta{}^r - \boldsymbol{\theta}_1 A_\Phi{}^r\}. \tag{8}$$

The superscript r denotes quantities valid only in the radiation zone defined by (4) and (5). R_0, θ, and Φ form a set of spherical co-ordinates with origin at 0. If there is only a single antenna (with center at 0) erected along the vertical axis z from which θ is measured, one has

$$A^r = z_1 A_z{}^r; \quad A_z{}^r = \frac{\Pi}{4\pi}\frac{e^{-j0R_0}}{R_0}\int I_z' e^{j\beta z'\cos\theta}dz'; \tag{9}$$

$$E_\theta{}^r = cB_\Phi{}^r = -j\omega A_\theta{}^r = j\omega A_z{}^r \sin\theta. \tag{10}$$

It is easily shown that the time-average Poynting vector in the radiation zone, $\bar{\bar{S}}^r$ is given by

$$\bar{\bar{S}}^r = (\omega\beta/2\Pi)(A_\theta{}^r A_\theta{}^{r*} + A_\phi{}^r A_\phi{}^{r*})R_1. \tag{11}$$

$\bar{\bar{S}}^r$ is in watts per meter squared and R_1 is a unit vector.

* Decimal classification: R125. Original manuscript received by the Institute, April 11, 1941.

† Cruft Laboratory, Harvard University, Cambridge, Massachusetts.

A convenient radiation function[1] $K_r^2(\theta, \Phi)$, referred to an arbitrary reference current I_r, may be defined as follows:

$$K_r^2(\theta, \Phi) = (4\pi R_0\beta/I_rI_r^*\Pi)^2(A_\theta^rA_\theta^{r*} + A_\Phi^rA_\Phi^{r*}). \quad (12)$$

Symbols with an asterisk are complex conjugates of the corresponding symbols without asterisks. The total power transferred over a time average from the moving charges within a great sphere of radius R_0 containing only the antenna array, to the moving charges in the rest of the universe outside the great sphere, is given by

$$\bar{\bar{T}} = (c\Pi I_rI_r^*/32\pi^2)\int_0^{2\pi}\int_0^{\pi} K_r^2(\theta, \Phi) \sin\theta d\theta d\Phi. \quad (13)$$

The quantity

$$R_r^e = 2\bar{\bar{T}}/I_rI_r^* = (15/2\pi)\int_0^{2\pi}\int_0^{\pi} K_r^2(\theta, \Phi) \sin\theta d\theta d\Phi \quad (14)$$

is the radiation resistance in ohms referred to the current I_r. (Use has been made of the numerical values of c and Π in obtaining (14).) The directivity[2] in a direction specified by θ_m, Φ_m may be defined to be

$$D = 30K_r^2(\theta, \Phi)/R_r^e. \quad (15)$$

In the simpler special case of a single antenna as described for (9), one has

$$K_r^2(\theta) = (4\pi R_0\beta/I_rI_r^*\Pi)(A_\theta^rA_\theta^{r*}); \quad (16)$$

$$R_r^e = 15\int_0^{\pi} K_r^2(\theta) \sin\theta d\theta; \quad (17)$$

$$D = 30K_r^2(\theta)/R_r^e. \quad (18)$$

The Field of a Conducting Thread

The current-amplitude distribution along an infinitely thin conducting thread ($a \doteq 0$) which is driven at the center is[3]

$$I_z = I_0 \sin(H - \beta|z|)/\sin H. \quad (19)$$

The abbreviation $H=\beta h$ is used. If, instead of being driven at the center, the thread is immersed in an electric field of uniform amplitude along its length, the current distribution is[3]

$$I_z = I_0\left\{\frac{\cos\beta z - \cos H}{1 - \cos H}\right\}. \quad (20)$$

The two distributions reduce to the same form in the important special case of half-wave units for which $H=\pi/2$. In this case

$$I_z = I_0 \cos\beta z \quad (21)$$

in both (19) and (20).

In all practical cases the radius a of the antenna differs sufficiently from zero to affect appreciably the current distribution, and through it the vector potential and the electromagnetic field.[3,4] The distributions (19) and (20) are the limiting forms of a more general case as the radius approaches zero and the conductivity becomes infinite. They are moderately good approximations for copper if the following inequalities are satisfied:

$$2\log(2h/a) \gg 1 \quad (22)$$

$$\sin H > 0. \quad (23)$$

Upon substituting the distribution function (19) in (9), and carrying out the integration, one obtains the following well-known relation for the distant field:

$$-j\omega A_\theta^r = E_\theta^r = cB_\Phi^r = j60I_0\frac{e^{-j\beta R_0}}{R_0}\left\{\frac{\cos(H\cos\theta) - \cos H}{\sin\theta\sin H}\right\}. \quad (24)$$

The reference current I_0 is the input current. Frequently the maximum amplitude defined by

$$I_m = I_0/\sin H \quad (25)$$

is used as the reference current. With this choice the radiation resistance referred to maximum current, viz., R_m^e, is not the input resistance unless I_m is also I_0. It is, rather, a fictitious resistance which, when multiplied by the square of the current at the reference point, gives the total power radiated.

The radiation function referred to the input current amplitude is

$$K_0^2(\theta) = 4\left\{\frac{\cos(H\cos\theta) - \cos H}{\sin\theta\sin H}\right\}^2. \quad (26)$$

The radiation resistance referred to the input current is obtained by a rather intricate direct integration using (17) and (26). It is expressed by the familiar formula

$$R_m^e = \frac{30}{\sin^2 H}\{-\cos 2H\,\overline{Ci}\,4H + 2(1+\cos 2H)\overline{Ci}\,2H + \sin 2H(Si\,4H - 2Si\,2H)\}. \quad (27)$$

Functions $Si\,x = \int_0^x \sin u/u\,du$, $\overline{Ci}\,x = \int_0^x (1-\cos u)/u\,du$ are tabulated in reference 4. $\overline{Ci}\,x$ can also be evaluated from $\overline{Ci}\,x = 0.5772 + \log x - Ci\,x$. It is perhaps well to repeat that (24), (26), and (27) are moderately good approximations for practical antennas that satisfy (22) and (23).

[1] The radiation function as here defined is closely related to a similar function due to S. A. Schelkunoff as reported by Donald Foster, "Radiation from rhombic antennas," Proc. I.R.E., vol. 25, p. 1329, footnote 5; October, 1937.

[2] This definition of directivity is equivalent to that of P. S. Carter, C. W. Hansell, and N. E. Lindenblad, "Development of directive transmitting antennas by R.C.A. Communications," Proc. I.R.E., vol. 19, p. 1802; October, 1931.

[3] E. Hallén, "Theoretical investigations into the transmitting and receiving qualities of antennas," *Nova Acta Upsaliensis*, ser. IV, vol. 11, no. 4, 1938.

[4] L. V. King, "On the radiation field of a perfectly conducting base insulated cylindrical antenna over a perfectly conducting plane earth, and the calculation of radiation resistance and reactance," *Phil. Trans. Royal Soc.*, London, vol. 236, pp. 381–422; November 2, 1937.

The condition (23) is actually satisfied by a large class of linear radiators commonly used as elements in composite arrays. Usually they fall within the range

$$0 < H < \pi. \tag{28}$$

The isolated half-wave element with

$$H = \pi/2 \tag{29}$$

is of particular importance. For it, one has

$$I_m = I_0; \qquad \sin H = 1 \tag{30}$$

$$-j\omega A_\theta{}^r = E_\theta{}^r = cB_\Phi{}^r$$

$$= j60I_m \frac{e^{-j\beta R_0}}{R_0} \frac{\cos\left(\frac{\pi}{2}\cos\theta\right)}{\sin\theta}; \tag{31}$$

$$K_m{}^2(\theta) = 4\left\{\frac{\cos^2\left(\frac{\pi}{2}\cos\theta\right)}{\sin^2\theta}\right\}; \tag{32}$$

$$R_m{}^e = 30\,\overline{Ci}\,2\pi = 73.13 \text{ ohms}. \tag{33}$$

It can be shown from the rigorous analysis of Hallen[3,5] that for practical antennas of cylindrical form for which the following limiting values obtain:

$$150 \leqq 2h/a \leqq 10^6, \tag{34}$$

the radiation resistance for $H=\pi/2$ falls within the range

$$58 \leqq R_0{}^e \leqq 67 \text{ ohms}. \tag{35}$$

The value 73.13 is obtained from the rigorous analysis in the limit as the radius a approaches zero.

The Equivalent Length of a Conducting Thread

Many composite antenna arrays are constructed of elements with H in the range specified by (28). In determining the distant field, the radiation function, the radiation resistance, and the directivity of such arrays, extremely intricate expressions are frequently encountered, especially if the elements are not parallel. In order to simplify the analysis, it is common practice[6] to replace the actual elements by "equivalent" ones with shorter so-called "effective lengths" and with uniform current-amplitude distributions. The conventional "effective length" $2h_e$ of such an "equivalent" element is defined by[7]

$$2I_0h_e = \int_{-h}^{+h} I_z dz. \tag{36}$$

(For an antenna of length h on a perfectly conducting plane one has

$$I_0h_e = \int_0^h I_z dz.)$$

[5] Ronold King, "The input impedance of a symmetrical antenna." Presented, American Physical Society, February 21, 1941.

[6] The curves in Fig. 383, p. 665, and Fig. 386, p. 667. of F. E. Terman, "Radio Engineering," McGraw-Hill Publishing Co., New York, N. Y., 1937, for example, are computed using this simplification.

[7] I.R.E. Standards on Transmitter and Antennas, 1938, p. 28.

For a center-driven unit with current distribution given by (19), the "effective length" as defined by (36) is obtained from

$$2I_0h_e = \frac{I_0}{\sin H}\int_{-h}^{+h} \sin(H - \beta|z|)dz. \tag{37}$$

Upon carrying out the integration, one obtains directly

$$\beta h_e \equiv H_e = \frac{1-\cos H}{\sin H} = \tan(H/2). \tag{38}$$

For a receiving antenna in a uniform electric field parallel to its axis and with the current distribution specified in (20), the "effective length" is calculated from

$$2I_0h_e = \frac{I_0}{1-\cos H}\int_{-h}^{+h} (\cos\beta z - \cos H)dz. \tag{39}$$

The integration leads to

$$\beta h_e = H_e = \frac{\sin H - H\cos H}{1 - \cos H}. \tag{40}$$

For $H=\pi/2$ both (38) and (40) reduce to the same expression.[8] It is

$$\beta h_e = H_e = 1. \tag{41}$$

The distant field of a center-driven element of length h_e and with a uniform current-amplitude distribution I_0 is readily obtained using (9). It is

$$-j\omega A_\theta{}^r = E_\theta{}^r = cB^r = j60I_0\frac{e^{-j\beta R_0}}{R_0}H_e\sin\theta. \tag{42}$$

The radiation function is

$$K^2(\theta) = 4H_e{}^2\sin^2\theta. \tag{43}$$

The radiation resistance is

$$R^e = 80H_e{}^2. \tag{44}$$

If the antenna is end-loaded and sufficiently short to satisfy the condition,

$$H^2 \ll 1, \tag{45}$$

the expressions (42) to (44) are correct with H written for H_e. If the antenna is not end-loaded, these expressions are more or less good approximations of (24), (26), and (27), respectively, depending upon the accuracy of a representation in terms of an "effective length" of antenna with uniform current-amplitude distribution. If the conventional definition of "effective length" as given in (37) and (38) is used, the approximation actually is not particularly good except for small values of H as defined by (45). In this case (38) simplifies to

$$h_e = h/2 \tag{46}$$

[8] In reference 7 the statement is made: "The effective height of a given antenna is the same for transmission as for reception in the usual case in which the current distributions are similar." This statement is vague and misleading.

and (42) to (44) become precisely the same as (24), (26), and (27), respectively, when only the leading terms are retained according to (45). For larger values of H, on the other hand, the situation is different. For example, with $H=\pi/2$, (38) leads to $H_e=1$ or $h_e=1/\beta=\lambda/2\pi$. Equations (42) to (44) then give

$$j\omega A_\theta{}^r = E_\theta{}^r = cB^r = j60I \sin \theta e^{-j\beta R_0}/R_0 \tag{47}$$

$$K^2(\theta) = 4 \sin^2 \theta \tag{48}$$

$$R^e = 80 \text{ ohms.} \tag{49}$$

These expressions do not agree particularly well with (31) to (33). Thus, in the most obvious case, there is an error of practically 10 per cent in the radiation resistance. The explanation for this large difference is not that given, for example, by Vilbig[9] who states that the forms (31) to (33) are based "on a more exact determination of the effective height which takes into account the not perfectly sinusoidal current distribution." Since (31) to (33) are calculated on the assumption of an *exactly* sinusoidal current distribution, this statement is erroneous. The large discrepancy between (47) to (49) and (31) to (33) is due to two things. In the first place, an antenna with a nonuniform current-amplitude distribution cannot be represented exactly by a fictitious antenna of different "effective" length with a physically impossible uniform distribution of current. In the second place, the definition of an "effective length" by (36) depends upon wishful thinking rather than upon any mathematical theorem to assure that the best possible approximation is obtained. Only in the case of a receiving antenna in a sensibly uniform electric field parallel to its axis is the choice of definition of "effective length" by (36) a really good one. In this case the time-average power received from a transmitter setting up a field E_z is given by

$$\bar{\bar{T}} = \int_{-h}^{+h} E_z I_z dz. \tag{50}$$

If E_z is essentially constant in amplitude over the length of the conductor (as is true in the distant zone), it may be removed from under the sign of integration. Then, with (36), one can write

$$\bar{\bar{T}} = E_z \int_{-h}^{+h} I_z dz = E_z 2h_e I_0. \tag{51}$$

Here h_e is properly defined to be an "effective length or height." No similar argument can be advanced for defining the "effective length" of a driven antenna by (36), nor does the reciprocity theorem apply.

The Effective Length of a Driven Antenna

In order to define an "effective length" h_e for a driven antenna which will make the simple expressions (42) to (44) the best possible approximations of the general forms (24), (26), and (27) for a sine distribution of current, one can make use of the method of Fourier. This provides an analytical mechanism for representing any function in terms of a series of simple trigonometric functions in a given interval. In the present instance, one wishes to represent the function

$$V(\theta) = \frac{\cos (H \cos \theta) - \cos H}{\sin H \sin \theta} \tag{52}$$

by the leading term in a simple sine series. That is, one wishes to write,

$$V(\theta) \doteq b_1 \sin \theta \tag{53}$$

where the coefficient b_1 is a differently defined H_e. The value of b_1 which involves the smallest error is specified by the familiar definition of the Fourier coefficients. The magnitude of the error involved is most readily estimated by evaluating the coefficient of the next term in the Fourier series representing (52). That is, one must determine both b_1 and b_3 in the series

$$V(\theta) = b_1 \sin \theta + b_3 \sin 3\theta + \cdots, \tag{54}$$

defined in the interval, $0 \leqq \theta \leqq \pi$. So long as b_1 satisfies the condition,

$$b_1 \gg b_3, \tag{55}$$

(53) is a good approximation of (52). (Cosine terms and even sine terms obviously have zero coefficients, so they have been omitted from (54).)

The coefficients b_n of the series (54) are defined by Fourier. They are

$$b_n = (2/\pi) \int_0^\pi \left\{ \frac{\cos (H \cos \theta) - \cos H}{\sin H \sin \theta} \right\} \sin n\theta d\theta. \tag{56}$$

Using the well-known relations,

$$\sin 3\theta = 3 \sin \theta - 4 \sin^3 \theta,^{10} \tag{57}$$

$$\pi/2 = \int_0^\pi \sin^2 \theta d\theta,^{10} \tag{58}$$

$$\pi J_0(H) = \int_0^\pi \cos (H \cos \theta) d\theta,^{11} \tag{59}$$

$$(\pi/H) J_1(H) = \int_0^\pi \cos (H \cos \theta) \sin^2 \theta d\theta,^{11} \tag{60}$$

one readily obtains

$$b_1 = 2[J_0(H) - \cos H]/\sin H \tag{61}$$

$$b_3 = 2[3J_0(H) - (4/H)J_1(H) - \cos H]/\sin H. \tag{62}$$

Subject to (55), the best "effective length" $h_e = H_e/\beta$ is obtained from

$$b_1 = H_e = 2[J_0(H) - \cos H]/\sin H. \tag{63}$$

[9] F. Vilbig, "Hochfrequenztechnik," Akademische Verlagsgesellschaft, Leipzig, Germany, 1937, p. 168.

[10] B. O. Peirce, "Table of Integrals," Ginn and Company, New York, N.Y., 1929.

[10] *Ibid.*, formula 261.

[11] N. W. McLachlan, "Bessel Functions for Engineers," Oxford University Press, New York, N.Y., 1934, p. 44, formula 22.

[11] *Ibid.*, p. 52, example 22.

For an antenna which is sufficiently short to satisfy (45), this reduces to

$$b_1 = H_e = 2(1 - H^2/4 + \cdots - 1 + H^2/2 - \cdots)/H$$
$$= H/2. \tag{64}$$

This agrees with (46). On the other hand, for $H=\pi/2$, (63) gives

$$b_1 = H_e = 0.945, \tag{65}$$

while (62) leads to

$$b_3 = -0.052. \tag{66}$$

Since (55) is moderately well satisfied ($b_3/b_1=0.055$),

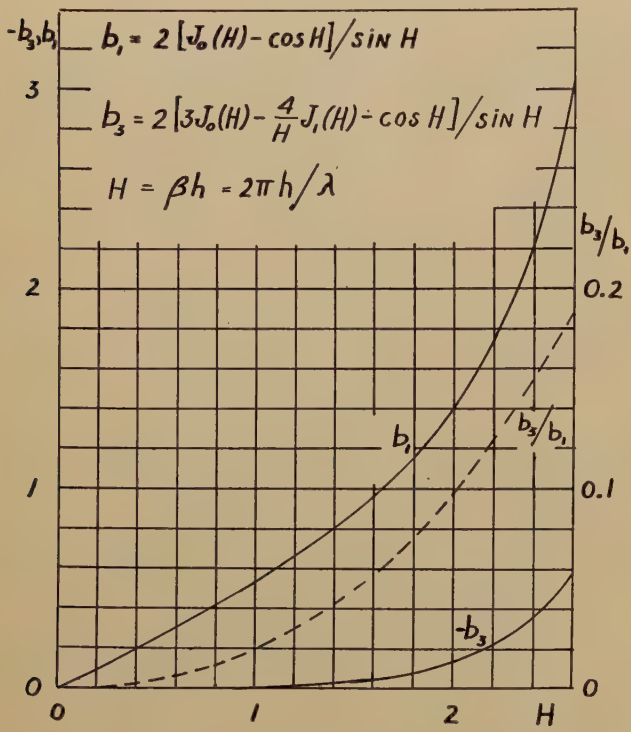


Fig. 1—Fourier coefficients b_1 and b_3 as functions of H.

one concludes that a representation in terms of sin θ alone is a good approximation. One then has

$$j\omega A_\theta^r = E_\theta^r = cB^r = j56.7I_0 \sin\theta \frac{e^{-j\beta R_0}}{R_0} \tag{67}$$
$$K_0^2(\theta) = 3.37 \sin^2\theta \tag{68}$$
$$R_0^e = 71.6 \text{ ohms.} \tag{69}$$

It is to be noted that these values are a very much better approximation of (31) to (33) than are (42) to (44). In particular, the radiation resistance differs from the correct value (33) for a sine distribution by only 2 per cent instead of 10. Moreover, since it is slightly less, rather than considerably more than this value, it is actually in even better agreement with the range of values of (35) for antennas of nonvanishing radius than is the exact value for an antenna of vanishing radius. Accordingly, the relations (67) to (69) are very satisfactory approximations for cylindrical antennas of small radius, whereas the formulas (47) to (49), which are derived from the conventional definition of "effective length," are poor approximations.

The functions b_1 and b_3 defined by (61) and (62), respectively, together with the ratio b_3/b_1, are plotted as functions of H in Fig. 1. It is clear that (55), and hence a representation using only the leading term in the Fourier series (54), is possible only for $H \leqq 2$. At $H=2$, the ratio b_3/b_1 becomes as large as 0.1. For larger ratios the second term in the series could not be neglected without appreciable error. One may conclude, therefore, that the general function (52) for the field of an antenna with a sinusoidal current distribution may be represented with reasonable accuracy by the simple sine function (53) in the range

$$0 \leq H \leq 2. \tag{70}$$

The field function $V(\theta)$, and all others derived from it, such as the radiation function and the radiation resistance, then have the same form as the corresponding functions (42) to (44) of a fictitious antenna of length h_e with a uniform current distribution. The length h_e, which may legitimately be called an "effective length," is related to the Fourier coefficient of the leading term in the series according to

$$h_e = H_e/\beta = b_1/\beta. \tag{71}$$

From the properties of the Fourier coefficients one knows that this is the best possible definition of an "effective length." Moreover, it is clear that a representation of the field function by the leading term in the series, and the definition of an "effective length," are useful only subject to (55) or (70).

The Distant Field of Two Crossed Antennas

An instructive example of the application of the simpler formulas obtained by using the leading term

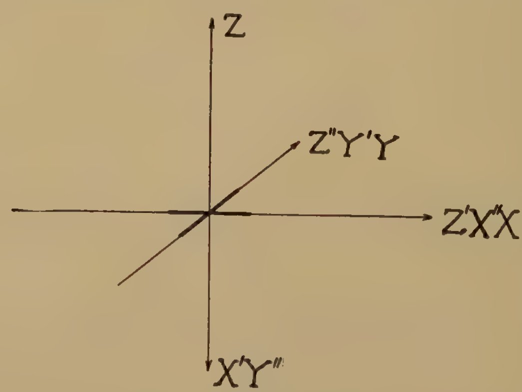


Fig. 2—Orientation of axes.

of the Fourier series is the case of two identical center-driven antennas crossed at right angles. This arrangement constitutes one element of the type used in the turnstile antenna.[12] In practice the individual antennas easily fall within the limits of length contained in (70), so that if the condition (22) is satisfied, one may express the field functions of the two units in the simple form (42) instead of (24) and readily combine

[12] G. H. Brown, "The turnstile antenna," *Electronics*, vol. 9, pp. 15–18; April, 1936.

them. Let antenna "prime" lie along the z axis of a primed system of co-ordinates shown in Fig. 2, while antenna "double-prime" lies along the z axis of a double-primed system. The vector potentials due to the two units are

$$A_{z'}{}' = FI_0{}'; \qquad A_{z''}{}'' = FI_0{}''. \tag{72}$$

Here

$$F = \frac{-60H_e}{\omega \sin H} \frac{e^{-j\beta R_0}}{R_0} \tag{73}$$

since with (10),

$$V(\theta')/\sin\theta' = V(\theta'')/\sin\theta'' = H_e. \tag{74}$$

If the two vector potentials are now referred to an unprimed set of rectangular co-ordinates as shown in Fig. 2, the resultant field of both units may be expressed in the following simple form:

$$A_x{}^r = FI_0{}', \quad A_y{}^r = FI_0{}'', \quad A_z{}^r = 0. \tag{75}$$

The spherical components of the vector potential, viz., $A_\theta{}^r$ and $A_R{}^r$, ($A_R{}^r$ does not contribute to the distant E and B fields, to the radiation function, or to the radiation resistance) are obtained using the standard formulas:

$$A_\theta = A_x \cos\theta \cos\Phi + A_y \cos\theta \sin\Phi \tag{76}$$

$$A_\Phi = A_y \cos\Phi - A_x \sin\Phi. \tag{77}$$

After writing

$$I_0{}'' = I_0 k e^{-j\delta}; \qquad I_0{}' = I_0 \tag{78}$$

and performing some simple algebraic manipulations, one readily obtains the following general formulas for the distant field

$$A_\theta{}^r = FI_0 \cos\theta(\cos\Phi + k\cos\delta\sin\Phi - jk\sin\delta\sin\Phi) \tag{79}$$

$$A^r = FI_0(k\cos\delta\cos\Phi - \sin\Phi - jk\cos\Phi\sin\delta). \tag{80}$$

$\boldsymbol{E}^r$ and $\boldsymbol{B}^r$ may be calculated from these expressions using (7) and (8). The radiation function as defined in (12) reduces to

$$K_0{}^2(\theta, \Phi) = \frac{4H_e{}^2}{\sin^2 H}\{\cos^2\theta(\cos^2\Phi + 2k\cos\delta\sin\Phi\cos\Phi + k^2\sin^2\Phi) + (k^2\cos^2\Phi - 2k\cos\delta\sin\Phi\cos\Phi + \sin^2\Phi)\}. \tag{81}$$

The radiation resistance defined by (14) is obtained by direct integration of (81). It is

$$R_0{}^e = 80H_e{}^2(1 + k^2)/\sin^2 H. \tag{82}$$

For half-wave units each carrying a current of the same amplitude, this simplifies to

$$R_0{}^e = 143.2 \text{ ohms}. \tag{83}$$

This is just double the value (69) obtained for a single unit.

The most impotant special case of the horizontal crossed antennas is that used in the turnstile array. Here,

$$k = 1; \qquad \delta = \pi/2. \tag{84}$$

In this case,

$$A_\theta{}^r = FI_0 \cos\theta e^{-j\Phi}; \qquad A_\Phi{}^r = -FI_0 e^{j\Phi}. \tag{85}$$

Along the horizontal plane containing the antennas, $\theta = \pi/2$, $A_\theta{}^r = 0$. The field is thus horizontally polarized, and its amplitude is independent of Φ. Subject to (84) the radiation function is

$$K_0{}^2(\theta, \Phi) = 4H_e{}^2(\cos^2\theta + 1)/\sin^2 H. \tag{86}$$

The radiation resistance is the same as (82). The value of $H_e = b_1$ may be taken from the curve of Fig. 1, using the given value of $H = \beta h$ for the antennas.

Other special cases of the crossed antennas derived from different phase and amplitude relations may be discussed as well as a stacked array forming the complete turnstile and many other arrays. The present purpose of illustrating the usefulness of the approximate formulas obtained from the leading term in the Fourier series has, however, been served. The mathematical simplification obtained by approximating (53) by (52) is clearly very great. Moreover, by using the Fourier coefficient rather than the conventional "effective length," one obtains not only the approximate shape of the field pattern of a composite array, but also good approximations of the magnitude of the radiation function, of the total power radiated, and of the radiation resistance for all antennas falling within the limits imposed by (22) and (70).

On the Energy Equation in Electronics at Ultra-High Frequencies*

C. K. JEN†, ASSOCIATE, I.R.E.

Summary—*A general equation for the rates of energy change in electronic devices at ultra-high frequencies is derived on the basis of classical electromagnetic theory, as an extension of a previous result. This provides the link between the directly observable quantities external to the electrodes and the dynamic quantities in the interelectrode space, together with radiations to remote regions. A part of these radiations is directly produced by space charges in accelerated motion. A qualitative discussion is given for the important role of the field propagation time within an electron-tube system at frequencies whose corresponding wavelengths are comparable with the tube dimensions.*

Introduction

In a recent paper by the author,[1] it was proved that the spatial quantities in an electron-tube system are related to the observable quantities on the electrodes through electrostatic induction. Emphasis was there laid on the necessity of revising the conventional concept of current, particularly at ultra-high frequencies. The derivation was limited to time-varying electrostatic fields (i.e., they could be functions of time but must always be derivable from scalar potentials) and also to infinite velocities for the propagation of fields. These assumptions are of course not rigorously true for general electromagnetic fields. It seems desirable to remove these limitations and at the same time point out some new aspects of the problem.

General Energy Equation

We assume, as before, N electrodes at specified potentials $\Phi_1, \Phi_2, \cdots, \Phi_N$ referred to a common ground potential. Each electrode is to be an equipotential at any instant. In the space the charge density ρ and the velocity V functions are supposed to be given and they always obey the fundamental equations of continuity and motion.

We now require that all quantities must satisfy Maxwell's equations in space and at the same time they must be consistent with our specified boundary values on the electrode surfaces. Writing only the relevant equations in the usual notation, we have

$$\frac{4\pi}{c}\rho \boldsymbol{V} + \frac{1}{c}\dot{\boldsymbol{E}} = \text{curl}\,\boldsymbol{H} \qquad \text{(1a)}$$

$$-\frac{1}{c}\dot{\boldsymbol{H}} = \text{curl}\,\boldsymbol{E} \qquad \text{(1b)}$$

and

$$E_k = (\boldsymbol{n}\boldsymbol{E})_k \qquad (2)$$

where E_k is the magnitude of the electric intensity at any point on the surface of the kth electrode and n is the unit normal drawn from the space to the electrode.

Multiplying scalarly (1a) by $\boldsymbol{E}$ and (1b) by $\boldsymbol{H}$ and subtracting the one product from the other ,we get

$$-\frac{c}{4\pi}\text{div}\,[\boldsymbol{E}\times\boldsymbol{H}] = \rho \boldsymbol{V}\boldsymbol{E} + \frac{1}{8\pi}\frac{d}{dt}(E^2+H^2) \qquad (3)$$

where $[\boldsymbol{E}\times\boldsymbol{H}]$ is the usual notation for the vector product of $\boldsymbol{E}$ and $\boldsymbol{H}$ (in the absence of this special sign, all multiplications are considered to be scalar). We now integrate (3) throughout a finite space, bounded on one hand by the single surface S consisting of all the electrode surfaces and their connections to a common zero-potential point and on the other a closed imaginary surface Ω enclosing the whole of S. Integrating in this way we have, by transforming the left-hand volume integral into a surface integral,

$$-\frac{c}{4\pi}\int[\boldsymbol{E}\times\boldsymbol{H}]_n dS - \frac{c}{4\pi}\int[\boldsymbol{E}\times\boldsymbol{H}]_n d\Omega = \int \rho \boldsymbol{V}\boldsymbol{E}d\tau + \frac{d}{dt}\frac{1}{8\pi}\int(E^2+H^2)d\tau. \qquad (4)$$

The left-hand integrals are the surface integrations of the well-known Poynting's vector $c[\boldsymbol{E}\times\boldsymbol{H}]/4\pi$. Each represents the rate of flow of energy through the corresponding closed surface.

Consider the first integral on the left-hand side of (4). Since $[\boldsymbol{E}\times\boldsymbol{H}]\perp\boldsymbol{E}$ and by (2) $\boldsymbol{E}$ is perpendicular to the electrode surface, therefore, $[\boldsymbol{E}\times\boldsymbol{H}]_n=0$. Thus the integration on the electrode surfaces vanishes everywhere. On the other hand, the integration for the portion of surface from any electrode to the common junction is in general not equal to zero. Consider any one of such surfaces, say the kth. Without loss of generality, we can consider this surface as being represented by a linear circuit element with a potential difference Φ_k between its terminals and a uniform current i_k flowing through (assuming that the capacitive effects are negligible for these "leads.") Let $dS = ds\,dl$, where ds is the infinitesimal length along the current and dl is the infinitesimal length along the periphery of the cross section perpendicular to ds. Since $\oint H_l dl = 4\pi i/c$ and $\Phi = -\int E_s ds$, we thus have for the kth contribution to the integral as $\Phi_k i_k$. The final result is accordingly

$$-\frac{c}{4\pi}\int[\boldsymbol{E}\times\boldsymbol{H}]_n dS = \sum_{k=1}^{N}\Phi_k i_k \qquad (5)$$

where i_k is the external current flowing to the kth electrode and according to the equation of continuity is related to the change of charge and the space current flowing to the electrode as follows:

* Decimal classification: R149. Original manuscript received by the Institute, April 18, 1941.

† Radio Research Institute, National Tsing Hua University, Kunming, China.

[1] C. K. Jen, "On the induced current and energy balance in electronics," Proc. I.R.E., vol. 29, pp. 345–349; June, 1941.

$$i_k = \dot{q}_k - \int_k (\rho V)_n dS. \qquad (6)$$

Combining (4) and (5), we finally get

$$\sum_{k=1}^{N} \Phi_k i_k = \int \rho V E d\tau + \frac{d}{dt} \frac{1}{8\pi} \int (E^2 + H^2) d\tau + \frac{c}{4\pi} \int [E \times H]_n d\Omega. \qquad (7)$$

Equation (7) represents a general equation of energy change in electronic devices. Let the surface Ω recede to infinity in the limit. We now interpret the term on the left-hand side of (7) as the net instantaneous power supplied to the space by all the sources external to the electrodes. The first term on the right-hand side is the rate of change of kinetic energy for all the space charges. The second term represents the rate of change of electromagnetic field energy. The last term is the power radiated out to infinite regions. It will be noted the result obtained in the previous paper[2] is a special case of this, with the difference that the change of magnetic-field energy and the radiation terms are neglected in the quasi-electrostatic treatment. The special case holds true for many purposes, even at ultra-high frequencies, but our present generalization adds certain other important features particularly concerning radiation.

Retarded Fields and Radiations due to Space- and Surface-Charge Distributions

The electric and magnetic fields in (7) are derivable from a scalar potential ϕ and a vector potential A according to the relations

$$E = -\operatorname{grad} \phi - \frac{1}{c}\dot{A} \qquad \operatorname{div} A = -\frac{1}{c}\frac{\partial \phi}{\partial t} \qquad (8)$$
$$H = \operatorname{curl} A$$

where the quantities ϕ and A must separately satisfy d'Alembert's differential equations if they are to be consistent with Maxwell's equations. It is well known that the solution for ϕ or A consists essentially of two parts, one due to the space distributions and the other due to surface distributions. We thus have

$$\phi = \int \frac{[\rho]}{r} d\tau + \sum_{k=1}^{N} \int \frac{[\sigma_k]}{r} dS_k \qquad (9a)$$

$$A = \frac{1}{c} \int \frac{[\rho V]}{r} d\tau + \sum_{k=1}^{N} \frac{1}{c} \int \frac{[i_k]}{r} ds_k \qquad (9b)$$

where

σ_k = surface-charge-density function on the kth electrode (related to q_k through the relation $q_k = \int \sigma_k dS_k$),

i_k = conduction current leading to the kth electrode, and

r = distance from the field to the point of integration. The quantities in the square brackets (this is not to be confused with the notation for vector products) must all be evaluated at a time $t - r/c$, while t is the instant at which the field is under consideration.

Equations (7), (8), and (9) give a consistent set of equations for the solution for our problem in electronics. A general solution for this is of course very complicated. However, one of the most interesting things is concerned with the possibility of radiation from the system. Ordinarily, an electron-tube system is not designed to radiate all by itself, the radiation being accomplished by an antenna network, which somehow absorbs power from the electron-tube system. This may be very accurately the case at low or moderately high frequencies. But at ultra-high frequencies, an electron-tube system must radiate energy to a certain extent. The radiation is brought about not only by the oscillating charges on conductors (as usually is the case) but also by the space charges in accelerated motion. It is on this account that the radiated energy (either intentional or unintentional) forms a non-negligible component in the balance of energies.

The theory of radiation from accelerated space charges is essentially the same as that of oscillating charges on conductors. The simplest and most typical example for this is that of a moving point charge, say an electron of charge e. If the electron velocity v is small as compared with c, we can obtain from (8) and simplified expressions for (9) the electric and magnetic field intensities at very large distances. They turn out to be $H = e[\dot{V} \times r]/r^2c^2$ and $E = [H \times r]/r$. The last term of (7) will then yield for a sphere of infinite radius $c\int[E \times H]_n d\Omega/4\pi = 2e^2\dot{v}^2/3c^3$, which is the classical result for the radiation of a single moving electron.[3] For a distribution of space charges with a spread which is much smaller than one oscillation wavelength the results are not much more complicated than a point charge and we can interpret the results on the same qualitative basis. If this is not the case, the situation is naturally very much complicated. In what follows we shall discuss only its qualitative bearing on our problem.

Field Propagation Time Within a Tube System at Quasi-Optical Frequencies

Let R = representative linear dimension of an electron-tube system, λ = oscillation wavelength, T_f = propagation time of a field variation across R, and T = period of oscillation. We have $R/\lambda = T_f/T$. Thus the above-mentioned comparability of R with λ means the same thing for T_f with T. This could be the case if frequencies higher than the present ultra-high range are used, temporarily designated here as the quasi-optical

[2] Equation (7) can be derived equally well from Green's theorem in much the same way as has been carried through in the previous paper, footnote 1, but the present derivation offers perhaps a simpler alternative.

[3] See for instance, W. Heitler, "The Quantum Theory of Radiation," Oxford Press, New York, N. Y., 1936, pp. 21 to 26.

frequencies. Under such conditions the signal of a field change in the space of the electron-tube-system could reach the electrodes (thereby calling forth a change in surface charges and hence also the external currents—the "induced" currents) with a time interval comparable with the period of oscillation. The same thing is true for the signal of a field change on the electrodes to reach charges in space. But according to (7) the instantaneous flow of currents together with the instantaneous values of potentials constitues an exact measure of the transfer of power between the electrodes and the space. Thus the amount of power transfer (net as well as instantaneous) is dependent on the field propagation times of the distributed space charges and field variations. As a part of the same balance of powers, the rate of radiation energy must also be dependent on the same factors. Hence the field propagation time is expected to play an important role in such considerations.

We recall that in the development of ultra-high frequencies, the transit time of electrons has been of great importance. Let T_e=electron transit time across R. Since $R_e/T_f=c/\overline{V}_e$ ($\overline{V}_e$=average velocity of an electron), T_e is generally very much larger than T_f for the usual manner of operation (in the order of 50 times at 400 volts). We see that $T \gg T_e \gg T_f$ is characteristic of low or moderately high frequencies and $T \sim T_e \gg T_f$ is characteristic of ultra-high frequencies. The negligibility of T_f has in these cases been very accurately true. The importance of T_e at ultra-high frequencies comes from the fact that there would be at any instant a "space-modulated" current, which introduces a rather harmful departure of phase for the conventional modes of operation. It is well known that numerous attempts have been made to remedy such an effect. For the forementioned quasi-optical frequencies the situation is $T_e \gg T_f \sim T$. This suggests the fact that, while the field propagation time is now very important, the electron transit time drops to a minor role. This paper, however, does not presume to go further than this suggestion.

High-Frequency Radio Transmission Conditions, July, 1941, with Predictions for October, 1941*

NATIONAL BUREAU OF STANDARDS, WASHINGTON, D. C.

THE radio transmission data herein are based on observations at Washington, D. C., of long-distance reception and of the ionosphere. Fig. 1 gives the July average values of maximum usable frequencies, for undisturbed days, for radio transmission by way of the regular layers of the ionosphere. The maximum usable frequencies were determined by the F layer at night and by the E, F_1, and F_2 layers during the day. Fig. 2 gives the expected values of the maximum usable frequencies for radio transmission by way of the regular layers, average for undisturbed days, for October, 1941. Average critical frequencies and virtual heights of the ionosphere layers as observed at Washington, D. C., during July are given in Fig. 3. Critical frequencies for each day of the month are given in Fig. 4.

Ionospheric storms are listed in Table I. The ionospheric storm beginning at 0100 E.S.T., July 5 was the

TABLE I

IONOSPHERIC STORMS

Day and hour E.S.T.	h_F before sunrise (km)	Minimum $f_F{}^0$ before sunrise (Mc)	Noon $f_{F_2}{}^0$ (Mc)	Magnetic character[1] 00–12 G.M.T.	Magnetic character[1] 12–24 G.M.T.	Ionospheric character[2]
July						
5 (from 0100)	[3]	[3]	[3]	6.6	6.5	7
6	420	1.6	4.7	3.1	4.1	6
7	346	1.8	<4.5	4.8	3.0	5
8 (through 1900)	325	2.7	5.2	2.6	2.6	3
10 (from 1400)	—	—	—	2.5	3.1	5
11	370	<1.6	<4.5	2.9	1.1	5
12 (through 0500)	332	1.9	—	2.0	2.1	3
16 (from 2200)	—	—	—	2.2	2.5	3
17 (through 1800)	316	1.8	<4.7	2.5	1.4	4
21 (from 0300)	308	1.6	<4.4	4.4	3.0	5
22 (through 1700)	328	2.0	5.4	2.8	2.8	3
For comparison: average for undisturbed days	296	2.92	5.78	1.9	1.9	1.2

[1] Average for 12 hours of American magnetic K figure determined by seven observatories, on an arbitrary scale of 0 to 9, 9 representing the most severe disturbance.

[2] An estimate of the ionospheric storminess at Washington, on an arbitrary scale of 0 to 9, 9 representing the greatest disturbance.

[3] No reflections observed above 1.7 Megacycles.

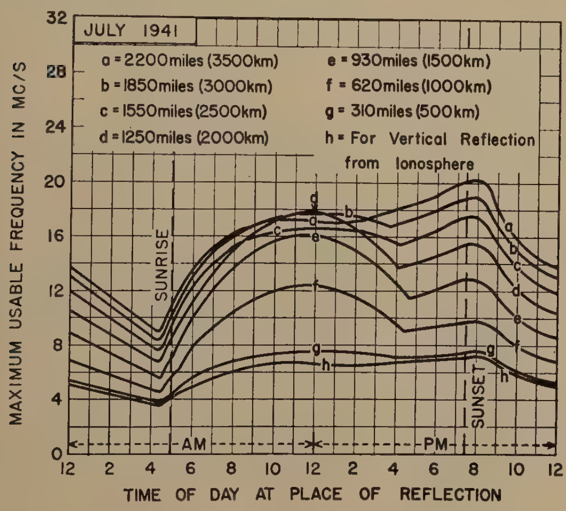


Fig. 1—Maximum usable frequencies for dependable radio transmission via the regular layers, average for undisturbed days, for July, 1941. The values shown were considerably exceeded during frequent irregular periods by reflections from clouds of sporadic E layer (see Table III). These curves and those of Fig. 2 also give skip distances, since the maximum usable frequency for a given distance is the frequency for which that distance is the skip distance.

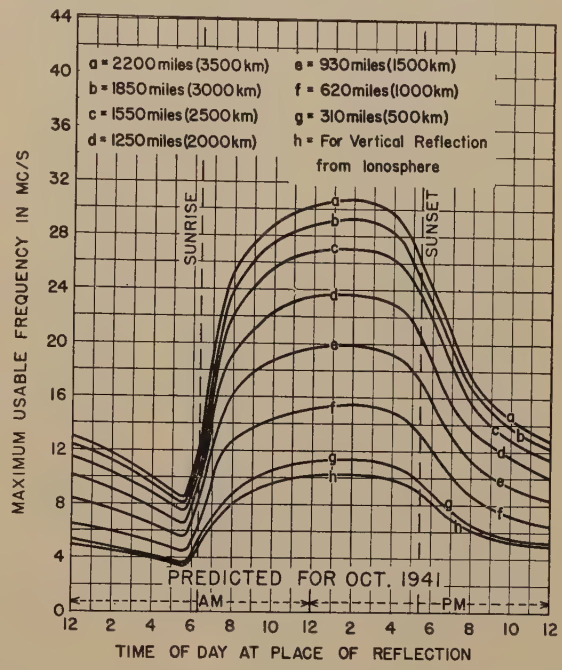


Fig. 2—Predicted maximum usable frequencies for dependable radio transmission via the regular layers, average for undisturbed days, for October, 1941. For information on use in practical radio transmission problems, see the pamphlets "Radio transmission and the ionosphere" and "Distance ranges of radio waves," obtainable from the National Bureau of Standards, Washington, D. C., on request.

TABLE II

SUDDEN IONOSPHERIC DISTURBANCES

Day	G.M.T. Beginning	G.M.T. End	Locations of transmitters[1]	Relative intensity at minimum[2]	Other phenomena
July					
1	2219	2250	Ontario	0.02	
3	1606	1940	Ohio, Ontario, D.C.	0.01	
5	2155	2225	Ohio, Ontario	0.0	
9	1323	1342	Ohio, Ontario, D.C.	0.0	Terr. Mag. Pulse 1322–1345
9	1420	1440	Ohio, Ontario, D.C.	0.02	
9	2255	2312	Ohio, Ontario, D.C.	0.01	
20	1757	1806	Ohio, Ontario, D.C.	0.02	
26	2105	2140	Ohio, Ontario, D.C.	0.0	Terr. Mag. Pulse,[3] 2105–2122

[1] W8XAL, Mason, Ohio, was not recorded regularly until July 10.

[2] Ratio of received field intensity during fade-out to average field intensity before and after, for station CFRX, 6070 kilocycles, 600 kilometers distant.

[3] As observed on the Cheltenham magnetogram of the United States Coast and Geodetic Survey.

* Decimal classification: R113.61. Original manuscript received by the Institute, August 11, 1941. Report prepared by N. Smith and C. O. Marsh.

most severe since that of March 30, 1941, and was characterized by a complete cessation of reflections between 0130 and 1500 E.S.T., due in part to a sharp decrease in ionization density and in part, especially

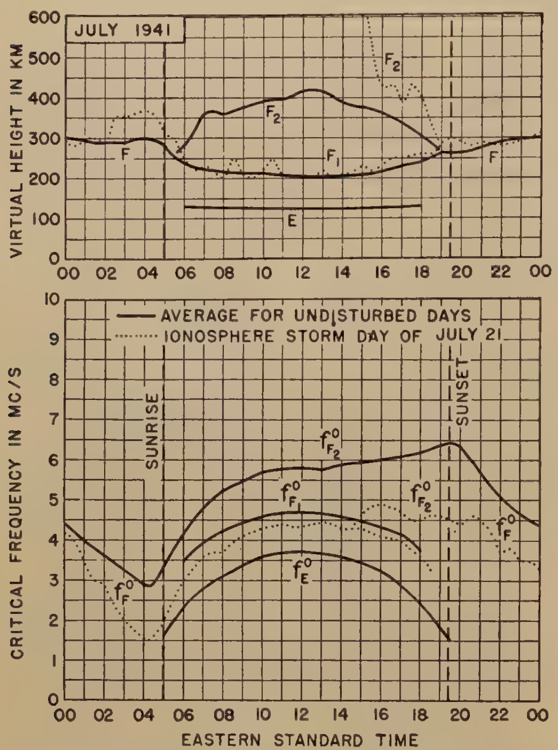


Fig. 3—Virtual heights and critical frequencies of the ionospheric layers, observed at Washington, D. C., July, 1941.

during the daytime, to a great increase in absorption in the lower ionosphere. The details of the ionospheric storm day of July 21 are shown in Fig. 3. The open circles in Fig. 4 indicate the noon and midnight critical frequencies observed during the ionospheric storms listed in Table I. The sizes of the circles roughly represent the severity of the storms.

Sudden ionospheric disturbances are listed in Table II. Table III gives the approximate maximum usable frequencies for good radio transmission via sporadic-E reflections.

TABLE III

APPROXIMATE MAXIMUM USABLE FREQUENCIES IN MEGACYCLES, FOR RADIO TRANSMISSION VIA STRONG SPORADIC-E REFLECTIONS

Day	Hour, E.S.T.																							
	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23
July 1																22	24	22		19	19	19	16	15
2	15																			23	17		23	23
3	23	23	26	33															21	21	18			19
4	25	34			40	21	24	32		37	34	38												
5		39	18																					43
6	44	23	23	25	23														45	33	32			
7																			31	18	50	50	34	23
8	18	17	15	16	19															24	40	15	20	
9																				18	34	34	24	
10																					20	16		
11	20	21	34	18	15				22	34		34	22	23	21			18			22	34	32	33
12			23	21			18											21	18	21	18	17	16	16
13	24	20	14	14	15																21	16	15	
14					16																16	23	19	17
15	29	36	40	18	21		19	22	23	35										16	19	18		
16						23	35					50	40	38	28			23	44					
17	25		22	22	23			20														22		
18						18	18	19	23	32	23	34	22	21		37	28	20	35					
19							20												18	18	23	17		
20								21	24							36	39	33	23	37				
21																			17					
22																34	20	23	18	21	19	19		
23				14	17	19	18												23	17				
24			16	18	18						21	23											20	17
25	15	17	17	15								23	23	22	23							23	28	23
26	18																				22	39	23	22
27	14															38					19	28	23	
29							18	18							50	35	24	50	50	50	34			
30																		16	19		28	18	20	23
31	24	15	13			19	18	21							22	25	33	19	19	50	38	21		

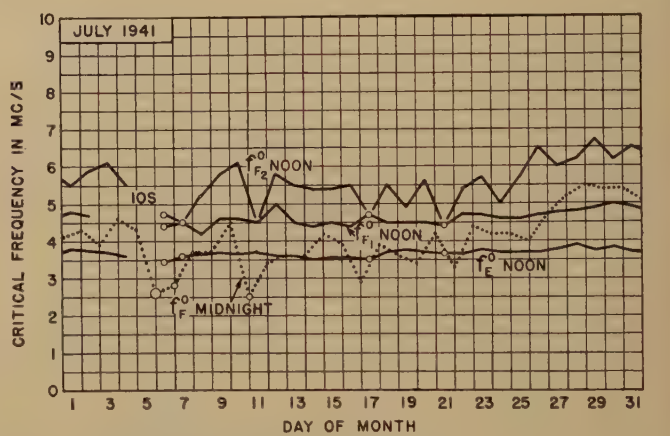


Fig. 4—Midnight and noon critical frequencies for each day of July. Open circles indicate critical frequencies observed during ionospheric storms. Sizes of circles represent approximately the severity of the storms. The letters "IOS" on July 5 indicate no reflections observed during a severe ionospheric storm.

Institute News and Radio Notes

EDITORIAL PROCEDURE

It is believed that the membership of the Institute will be interested in learning of the necessarily somewhat complicated normal procedure through which it is necessary that every paper shall pass while under consideration for publication in the PROCEEDINGS. These steps have been developed from experience through the years and are intended to ensure a high standard of papers and at the same time provide reasonable speed of editing and publication.

When the paper is received by the office of the Institute, its receipt is first acknowledged to the author, to whom a number of relevant questions as to previous nonpublication and the like are then addressed.

The subject matter of the paper then determines the selection of a particular member of the Papers Committee active in that field who is designated to read the paper and comment on it. The reader fills out an extensive and searching form concerning the paper.

The paper is also referred to a member of the Board of Editors who is regarded as especially competent in the field treated by the paper. The Editorial reader is required to consider not only the suitability of the paper for publication but also to suggest in detail all requisite editorial changes.

In the event that the author initially submits three copies of the paper, the paper can be considered simultaneously by three members of the Papers Committee, thus assuring that any abnormal delay by one of these reviewers will not prevent early initiation of the next editorial process. About sixty members serve on our Papers Committee and Board of Editors, contributing their technical abilities and time to a sound publication. Under the exigencies of business they cannot always give as prompt consideration to a paper as might be desirable.

In the case that serious or major criticisms of the paper are received, additional members of the Papers Committee and of the Board of Editors may be asked similarly to comment on the paper. It is the determined aim of the Institute editorial procedure to avoid any arbitrary or unduly limited judgment of the quality or contents of a submitted paper.

After the readings of the paper have then been completed, it is submitted to the Editor, the Co-ordinating Committee, or both. The Co-ordinating Committee is a numerically limited group, of circulating membership drawn from the Board of Editors to ensure reasonable policy uniformity in the decisions on the acceptability of all submitted manuscripts. Final approvals for publication originate either from the Co-ordinating Committee or the Editor. In every instance, the author of a submitted paper is given a clear analysis of his paper as it appears to the readers, and is afforded an opportunity to modify his paper along appropriately suggested lines if it appears that it may thus become acceptable for publication.

Accepted papers then pass to the assistant editor who puts the manuscript into proper condition to be sent to the printer. From this point, the setting up of galley proof, the reading of such proof, the preparation of page proof, and the reading of page proof follow as routine steps.

From the foregoing, authors will gather that the prompt and effective handling of manuscripts is furthered by carrying out the following suggestions. The length of the paper should be kept at the minimum consistent with an exposition of the subject matter with clarity. The illustrations should be appropriate for reproduction (information on this point being obtainable from the Institute). Three copies of the paper complete with illustrations should preferably be furnished. The original illustrations may be held by the author until the paper has been accepted for publication, so avoiding extensive handling and the dangers of successive mailings to the editorial reviewers. Care should be taken not to offer the same material for publication elsewhere prior to its appearance in the PROCEEDINGS. Any other publication plans should be stated by the author in his letter of transmittal of the manuscript.

The high standards of the PROCEEDINGS are the necessary consequence of the painstaking routine which has been described above. Accordingly it is hoped that it will be of interest to prospective authors and others.

Alfred N. Goldsmith, Editor

COMING MEETING

Rochester Fall Meeting
November 10, 11, and 12, 1941

Board of Directors

A special meeting of the Board of Directors was held in the Hotel Statler, Detroit, Michigan, on June 24, 1941. Those present were F. E. Terman, president; Haraden Pratt, treasurer; Austin Bailey, A. B. Chamberlain, I. S. Coggeshall, Virgil M. Graham, O. B. Hanson, R. A. Heising, L. C. F. Horle, F. B. Llewellyn, B. J. Thompson, H. M. Turner, A. F. Van Dyck, H. A. Wheeler, L. P. Wheeler, and H. P. Westman, secretary.

Action was taken at a previous meeting, approving the formation of a Twin Cities section on the completion of certain minimum requirements. The secretary

reported these requirements had been met and that the section was now established and in operation.

It was reported that the Rochester Fall Meeting would be held on November 10, 11, and 12, 1941.

The Secretary was instructed to have ballots mailed to the voting members of the Institute on several proposed changes in the Constitution.

A statement concerning the Institute's participation in regulatory and legislative matters in the United States was prepared to be read by President Terman at the banquet.

Harold R. Zeamans was appointed General Counsel to the Institute.

Sections

Atlanta

The meeting was devoted to a tour of the State Police radio facilities under the guidance of Captain Harris and Mr. Simmons of the Georgia State Patrol.

Captain Harris demonstrated the use of the fixed station transmitter. A general explanation of the operation of the system was presented. The meeting then adjourned to a parking lot where the mobile transmitter was located. The 25-watt low-frequency composite transmitters used for this purpose were then demonstrated. Reliable operation of these transmitters was stated to be ten miles and no dead spots are evident within the operating area of a mobile unit. The system provides two-way communication between the moble station and the fixed station.

April 18, 1941, A. W. Shropshire, chairman, presiding.

"Broadcast Antenna Measurements" was the subject of a paper by Ben Akerman, chief engineer of WGST. The desirability of such measurements was first considered. Four methods were discussed and included the use of a radio-frequency bridge, resistance-variation, series-resonance, and the voltmeter-ammeter methods. These various methods were outlined in detail and compared.

The remainder of the paper was devoted to a discussion of the radio-frequency bridge method. Some sources of errors encountered in the bridge method were listed as poor shielding, long leads, and false nulls.

Descriptions were given of a portable radio receiver and a radio-frequency oscillator which are used in making antenna resistance measurements. The receiver has had the automatic-volume-control action removed. The oscillator must possess sufficient stability to be unaffected by changes in load caused by the radio-frequency bridge.

Antenna resistance curves were shown. Numerous problems encountered in making the measurements from which the curves were drawn were discussed. In one case, a large percentage change in antenna resistance was noticed when the tower-lighting chokes were changed.

June 20, 1941, A. W. Shropshire, chairman, presiding.

Boston

H. J. Brown, president of the Engineering Societies of New England, discussed briefly the broad activities of that organization and the advantages which they bring to the Boston Section of the Institute which is a member of it. Among its activities are the work of committees dealing with the controversial question of engineering registration and the problem of providing an interesting program for the younger engineers fresh from their schooling.

"Atoms, Cyclotrons, and Radio" was the subject of a group of presentations by M. S. Livingston and J. H. Cook of Massachusetts Institute of Technology and J. J. Livingood and R. W. Hickman of Harvard University.

Professor Livingston described the type of reaction which goes on within the exhausted chamber of the cyclotron. He listed the rather extensive number of different kinds of particles and types of radiation which are involved. He then gave a few examples of the new types of chemical reactions which described successive stages of the process by which a cyclotron accomplishes the transmutation of elements. These equations require an additional type of notation in which subscripts before the chemical symbol denote the electric charge and superscripts following the symbol indicate the mass of each particle.

Professor Livingood showed his animated moving pictures which illustrate very clearly for the uninitiated the method of operation of the cyclotron and the mechanism of the reactions of transmutation.

Professor Hickman described the radio-frequency circuits which provide the 35 kilowatts of power used in the cyclotron at Harvard. The radio-frequency energy is fed to the dees over parallel lines. This method requires that the cantilever-type supports of the dees be insulated from ground. The comparatively high power used and the small distances involved make the problem of insulation not a simple one.

Mr. Cook described the radio-frequency circuit arrangements of the cyclotron at Massachusetts Institute of Technology. Here somewhat more power is used than at Harvard, and it is fed to the dees over large concentric lines. This obviates the necessity of using insulators at the accelerating-electrode supports as are required with parallel feed lines since the circuits may be conductively grounded at a node.

After the meeting everyone was invited to inspect the cyclotron at Massachusetts Institute of Technology.

October 25, 1940, W. L. Barrow, chairman, presiding.

A. B. Chamberlain, chief engineer of the Columbia Broadcasting System, presented the paper on "CBS International Broadcast Facilities" which he presented at the Washington Section. This paper was summarized in the May, 1941, issue of the PROCEEDINGS.

March 28, 1941, W. L. Barrow, chairman, presiding.

A. W. Friend of Cruft Laboratory and Blue Hill Meteorological observatory of Harvard University, presented a paper on "Tropospheric Reflections."

The results of tests made in co-operation with the United States Weather Bureau using records obtained with the new Simmonds-Lange radiometeorograph at East Boston Airport were compared with simultaneous tropospheric radio-wave-echo records taken at East Lexington, Massachusetts. It was shown that a calculated plot of the rate of change of dielectric constant of the air versus altitude closely matched the oscillographic pattern of radio-pulse-echo amplitude versus time.

The heights of even, very thin, cloud formations and the base of the stratosphere were measured quite accurately by the radio-wave echoes. Aircraft, meteorological, and wave-propagation-prediction applications were proposed.

J. A. Pierce of Cruft Laboratory, Harvard University, described the "Harvard Ionosphere Eclipse Expedition to South Africa." The field work and the preliminary results of the expedition to Queenstown, South Africa, were discussed. Observations of the ionospheric effects were also carried out by expeditions from the Australian Radio Research Board and the Bernard Price Institute of Geophysics of the University of the Witwatersrand of Johannesburg.

In general, the results of previous expeditions were confirmed, although the changes in the ion density of the F_2 layer were of a larger magnitude than had been reported before. Evidence was obtained of the differences in the intensity of the ionizing radiation from various parts of the surface of the sun.

May 2, 1941, W. L. Barrow, chairman, presiding.

Buenos Aires

P. J. Noizeux, chairman of the Buenos Aires Section, described his "Impressions of a Recent Trip to the United States."

The recent technical progress in the United States in the fields of aviation, television, frequency modulation, and communications in general were discussed. The technique employed in airport control towers for supervising the movement of many airplanes was described. Special reference was made to the air traffic between New York City and Washington, D. C.

The present status and public acceptance of frequency modulation were described. Brief comments were then given on a demonstration of color television and on the applications of television.

June 11, 1941, P. J. Noizeux, chairman, presiding.

Cleveland

"Vibrators and Vibrator-Operated Power Supplies" was the subject of a paper by Allen Nace of the Radiart Corporation.

The first "all-electric" automobile radio receivers came into use in 1932 when vibrator power supplies made them possible.

Poor voltage regulation was a problem with the early types and the methods used for improving this situation were described. By the use of suitable circuits, which reduce the residual magnetism in the core of the transformer, almost unlimited power output can be obtained.

The vibrator is essentially a low-power device, useful for powers up to 50 watts. Manufacturers recommend that for maximum life they should not deliver more than 18 watts continuously. For practical purposes, they are limited to operation at temperatures not exceeding 300 degrees Fahrenheit. It was stated that vibrators are more easily serviced than motor generators and are less expensive.

Synchronous vibrators, their circuits, advantages, and breakdown problems were discussed. Commutating vibrators were mentioned and reasons given as to why they are not generally used.

Tungsten has been found to be the only practical material for vibrator contacts.

As evidence of their practicability, it was stated that some of the airlines are beginning to accept vibrator power supplies.

Mary 22, 1941, C. E. Smith, chairman, presiding.

Connecticut Valley

"Frequency Modulation for Communication Purposes" was the subject of a paper by Fred Budelman, chief engineer of the F. M. Link Company. A brief history of the Connecticut State Police radio system was first given.

The phase-shift oscillator with crystal control proved effective for the transmitters. The use of class C operation throughout permits the filament emission to drop to 25 per cent of normal in most cases before any considerable difficulty is encountered. This simplifies service problems.

The receiver is more complicated than one for amplitude modulation. It provides higher gain, requires a limiter, and uses more tubes. The limiter contributes definitely to the stability of operation.

The car antennas are located on top of the car roof. They provide a circular field-strength pattern whereas if mounted at the read of the car a power gain of 10 to 1 would result. Such a pattern would be undesirable.

When signals are being received at low levels, the signal-to-noise ratio for frequency modulation is comparable with that for amplitude modulation. The noise level is improved when the signal is strong enough to result in limiter action.

Band width and stability requirements dictated a deviation frequency of 15 kilocycles. The male voice contains most of its important components in a frequency band below 3 kilocycles and by cutting off all upper frequencies, a deviation ratio of 5 may be used. This is about the same as is used for broadcast services.

Phase-shift modulation is basically a pre-emphasis system which reduces noise to a surprising degree. The narrower receiver band width, 40 kilocycles, will permit a weaker signal to suppress the noise. This is largely controlled by tolerances in design.

The advantages of frequency modulation in military services was then discussed. Because the stronger of two signals blankets the weaker, if few channels are available in a small area, the stronger signal will always be received. Thus, the interference of the enemy will be reduced. The transmitter power may be reduced till the signal is satisfactory at the place where reception is desired but falls off so rapidly beyond that an intelligible signal cannot be picked up. This has the added advantage of reducing interdivision interference and would permit a superior officer to control the reception to several divisions simultaneously by using a transmitter of far greater power than that used by any other station. The ignition interference which predominates in motor vehicles is much more easily eliminated in frequency modulation than in amplitude modulation. When employed in the high-frequency region, critical adjustments do not result. Transmitter design is considerably simplified.

This was the annual meeting of the Section and in the election of officers F. G. Webber of the F. W. Sickles Company was named chairman; W. M. Smith of the F. W. Sickles Company was elected vice chairman; and C. I. Bradford of the Remington Arms Company, Inc., became secretary-treasurer.

May 6, 1941, K. A. McLeod, chairman, presiding.

Dallas-Fort Worth

A paper on "Relays" was presented by J. E. Mossman, sales engineer for C. P. Clair and Company.

Relays were divided into five general classes depending on the size of the load they control. Supersensitive relays control currents of microampere magnitude, sensitive types have contacts to carry up to 5 to 10 amperes of alternating current, light-duty power relays handle around 15 to 20 amperes, the heavy-duty types operate with loads from 20 to 50 amperes, and contactors are used for heavy loads such as large motors.

The basic parts of a relay are the coil, magnetic structure, and the contacts. The size and shape of the magnetic structure governs the timing and quick-acting relays have long slim structures. The type of iron used is important and some information was given on annealing processes.

The size of the contacts do not necessarily indicate the current-carrying capacity. Consideration must be given to the use to which the relay will be put. The amount of contact area is not as important as the amount of metal attached to the contacts to carry away the heat. A wiping action is not as important as having the proper amount of pressure.

If was pointed out that the teletype machine has stimulated the development of relays. Relays having as many as twenty-two springs are used in these machines and operate more in a month than a dial-phone relay does in two years.

The need for a satisfactory relay for use with vacuum tubes has encouraged a great deal of competition. Whereas only ten manufacturers were in existence in the United States fifteen years ago, there are now eighty-seven companies making relays.

The use of 400-cycle power supplies in aircraft has made necessary the designing of new relays.

Numerous applications of relays were then described. The paper was concluded with a brief history of the Leach relay which was developed as a keying relay for radio transmitters.

During the business part of the meeting, the Constitution for Sections was adopted and the annual meeting of the Section was designated to occur during the month of December.

June 19, 1941, D. A. Peterson, chairman, presiding.

Emporium

A. L. Schoen of the Eastman Kodak Company, presented a paper on "The Electron Microscope."

The subject was introduced with a description of an optical microscope. The theory of resolution of lens systems was considered especially in regard to the resolution that could be expected from using different wavelengths of light. The work of de Broglie, Davisson, and Germer on the theory of particle wavelengths and particularly on the electron wavelengths for various voltages was mentioned.

There was then presented a description of the electron microscope designed and built in the Eastman Kodak Laboratories. The component parts, such as the lens system, plate holder, and voltage regulators, were clearly shown.

Following the discussion of the instrument, numerous slides of germs and crystals were shown. Since the Eastman Kodak Company is primarily interested in silver salt crystals and their changes during development, this phase was stressed. The photographs showed the changes in the crystals during development and illustrated clearly the formation of a "seaweed" structure after development. The effects of different developers and different times of development and the resulting change in crystal shape were shown.

It is anticipated that the use of the electron microscope in the study of germs, cells, and crystals will bring about outstanding discoveries in these fields.

The meeting was closed with the showing of a motion-picture film entitled "This Changing World" which was furnished by the International Nickel Company.

June 27, 1941, R. K. Gessford, chairman, presiding.

Four technical papers were presented at the Fifth Annual Summer Seminar which was held on August 1 and 2.

"Extending the Range of Q-Meter Measurements to Higher Frequencies" was the subject of a paper by C. J. Franks of the Boonton Radio Corporation. A brief review of the Q meter was first presented. The ordinary meter is limited to measurements below 30 megacycles because of high-frequency-resistance effects.

A new instrument has been designed for accurate measurements at frequencies up to 150 megacycles. The design features were described in detail. The many ingenious methods used to extend the range of the instrument illustrated effectively the differences in technique required for ultra-high-frequency work as contrasted with the lower frequencies.

L. E. Packard of the General Radio Company, gave a paper on "Impedance Measurements Over a Wide Frequency Range."

A brief review was first presented of the methods used for measuring impedance. In some work being done by the author, the dielectric properties of materials are of particular interest. A susceptance-variation circuit was described and its mechanical features discussed in detail. The instrument operates at frequencies between 0.5 megacycle and 100 megacycles.

Curves were shown of typical measurements on materials such as phenol fiber, phenol fabric, hard rubber, mycalex, and plasticized polystyrene. Graphs were also given of the dissipation factor versus frequency and of the capacitance versus frequency for the same materials.

A study of the dielectric properties of matter provides information about the atomic character of the substance and on the basis of these data, a theory can be formulated to explain the behavior of the material at low and at high frequencies.

"The Megacycle Meter" was described by Jerry Minter of Measurements Corporation. A grid-dip meter is a low-power oscillator with a meter in the grid circuit to indicate the strength of oscillation. The *LC* meter now on the market is a form of this instrument.

The megacycle meter is a form of grid-dip meter especially designed for operation between 10 megacycles and 400 megacycles. The mechanical features of the particular design were described. Its compact form and methods used to obtain continuous coverage of the frequency band were stressed.

The paper was closed with a discussion of the uses of the instrument which included adjusting and aligning ultra-high-frequency superheterodyne receivers and the locating of parasitic oscillations in ultra-high-frequency circuits.

M. A. Acheson, of the Hygrade Sylvania Corporation, presented a paper on "High-Frequency Tube Phenomena."

Some observations of tubes now in use were made first. In a brief review, some of the potential and current effects in the ordinary type of tubes were described. The capacitance effects in tubes with and without electron-current flow were considered. At high frequencies, the transit time of the electrons become important and a phase shift results.

The radiation of power and its relation to tube-element configurations and current flow as well as to frequency was discussed. It was pointed out that reduction in size of elements is not the most effective solution for high-frequency tubes but proper loading and coupling provides the answer.

The advantages of the lock-in tube for ultra-high-frequency work were pointed out. In addition to the element design, proper consideration must also be given to the manufacturing aspects.

In balancing all of the factors in ultra-high-frequency tubes, it is found that some design features that are not usually used must be incorporated in tubes while others which are usually observed are eliminated.

During these meetings, two sound motion pictures were projected. One "The Hurricane's Challenge" described the restoration of telephone service after the 1938 New England hurricane, and the other on "Nickel Highlights" showed the processes used in separating pure nickel from the ore as well as the treatment of the metal in the forming of finished products. The films were provided through the courtesy of the Bell System and the International Nickel Company, respectively.

Indianapolis

"The General Engineering Aspects in Regard to the Production of "'Fantasia'" was the subject of a series of papers by M. C. Batsel, chief engineer, and J. E. Volkman, J. L. Underhill, and P. W. Wildow, engineers of the RCA Manufacturing Company (Indianapolis).

The desire to take full advantage of the possibilities in the pickup and amplification by electrical means of orchestral music led Leopold Stokowski to collaborate with Walt Disney on the production of a "short," "The Sorcerer's Apprentice." Before it was completed, it was decided to enlarge the idea into a full-length feature picture. The pictorial effects and music were designed to complement each other. The RCA Manufacturing Company handled the recording, which was done by the variable-area method, and designed the special recording, printing, and projection equipment required.

Mr. Volkman, then discussed the "Acoustical Problems" The music was recorded at the Philadelphia Academy of Music. Five microphone channels were used for recording or controlling the different sections of the orchestra, such as the violins and violas, the brass, the percussion, the wood winds, and the bass. The microphone placement and group separations used to prevent overlapping were described.

The control channels and amplifiers for reproduction were then considered. A special loud-speaker system is used in the presentation of the picture. Three speaker assemblies are used on the stage, each assembly combining high-frequency and bass-response speakers in proper proportion. The assemblies are located at the left, center, and right of the stage to provide projection of special sound effects from the proper angles. In addition, a large number of loud speakers are arranged throughout the theater for added special effects and are controlled by the left and right channels respectively.

"The Optical Problems" were described by J. L. Underhill. To provide the greatest latitude in sound projection and control, a separate film was used to carry the sound. This required a synchronized drive between the sound and picture films. Four sound tracks were used, each of which was more than twice as wide as the standard. This permitted the use of so-called "push-pull" tracks, photocells, and amplifiers.

Three of the sound tracks were used for the three acoustic channels and the fourth was used for automatic gain control of the other three. For gain control, three different audio frequencies were used with suitable filters and rectifiers which provided grid bias for the respective amplifiers.

A special printer was developed to permit positioning of each sound track in its proper place on the film and in synchronization with the other sound tracks. The optical system of the printer provided an expansion of slightly over 2 to 1 in the width of the track across the film while maintaining a 1-to-1 ratio along its length. In the production of this 4-track film, use was made of cylindrical lenses to provide a slit of light. The surfaces of the lenses were coated to improve their transmission of light and the sharpness of the images throughout the system. The images of the sound tracks are spread horizontally by means of a lens system which projects them on cylindrical lenses which concentrate them on dual-plate photoelectric cells. From there on, the amplification and control is accomplished electrically.

In the presentation of this feature picture, two special sound projectors, two special film projectors, and nine complete racks of equipment are required in addition to the great number of special speakers. This accounts for the lack of duplicate showings in many cities and the adoption of a "road-show" policy of presentation.

P. W. Wildow presented "The Problems of Electrical Design." He discussed briefly the electrical circuits involved in the apparatus described by the other speakers and enlarged upon their observations.

Prior to the meeting, an inspection trip was taken through the RCA Manufacturing Company record-processing plant. The complete manufacture of phonograph records was observed. Although the original master record comes from a recording studio, all other "negative" records are processed in this plant and used for moulding purposes. The mixing of the compound, the pressing into final form, inspection and test, and shipping were discussed.

June 20, 1941, A. N. Curtiss, chairman, presiding.

Los Angeles

This was the annual amateur meeting and various speakers were introduced by D. C. Wallace. "Amateur Radio Preparation for National Defense" was discussed by Ralph Click, W6MQM and section communication manager for the American Radio Relay League. The Amateur Emergency Corps which was formed by the American Radio Relay League has resulted in the organization of a series of networks throughout Southern California. This work was outlined.

Don Shugg, W6ETJ, of Western Air Express, discussed "The Army Network." He outlined the work that is done in the handling of traffic at a maximum speed. It was pointed out that these nets operate principally on continuous-wave

transmission but lately have expanded into the phone bands.

"Ultra-High-Frequency in Air Transports" was the subject of Earl Kiernan of the Bendix Aviation Company. The work of amateurs and their contribution to the successful application of many high-frequency circuits was outlined.

W. A. Adams, W6ANN, has been interested in "Five-Meter DX." Located at Palos Verdes Hills, he has "worked" twenty-one states on five meters. He described various receiving systems and discussed the correlation of temperature-inversion weather conditions and radio wave propagation.

"Army Aircraft Radio" was presented by Vearn Warne of the Radio-Television Supply Company. He outlined the different systems being used and in construction in the vicinity of Southern California.

May 27, 1941, J. N. A. Hawkins, vice chairman, presiding.

The section was invited to the annual dinner meeting of the Founder Engineering Societies which was sponsored by the Los Angeles Engineering Council. At it a paper on "Science and the Rational Animal" was presented by Max Mason, chairman of the Observatory Council of the California Institute of Technology.

June 5, 1941.

Portland

"The Portland Communication System" was the subject of a paper by T. V. Ehmsen, radio technician for the Portland Police Department.

He discussed the development of the Portland Police Department communication system and described the features of the present equipment. A visit was then made to the dispatchers control room, the radio shop where most of the equipment is constructed, and to the transmitter which is also one of the three remote receiving locations.

June 26, 1941, E. R. Meissner, chairman, presiding.

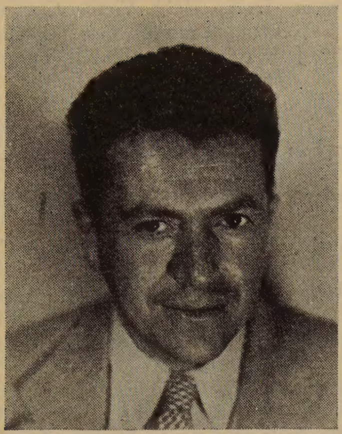
Membership

The following admissions to Associate grade were approved by the Board of Directors.

Adams, I. D., 822 N. Mills St., Orlando, Fla.
Alley, C. L., 806 S. Church St., Belleville, Ill.
Ayres, W. R., 415 S. Charles, Belleville, Ill.
Balog, M., 22 W. Fifth St., Emporium, Pa.
Basu, S. K., 92 Upper Circular Rd., Calcutta, India
Beers, R. F., Box M, Lincoln, Mass.
Bellamy, R. S., 2063 Stanley St., Montreal, Que., Canada
Boyd, D. L., The General Industries Company, Elyria, Ohio
Brown, R. A., 911 Oneida St., Joliet, Ill.
Bullock, R. E., 1827 Campus Rd., Los Angeles, Calif.
Cambre, A. C., Monroe 4951, Buenos Aires, Argentina
Combie, A. A., Massachusetts State Police, Radio Division. WMP, Framingham, Mass.
Corenthal, R., 2 Hinckley Pl., Brooklyn, N. Y.
Cox, F., 39 Warwick Crescent, Hayes, Middx., England
Dachin, N. V., Guemes Torino 84, Vicente Lopez, Buenos Aires, Argentina
Dutt, N. L., 33/2 Beadon St., Calcutta, India
Eckner, W. W., 15 Garden St., Ridgefield Park, N. J.
Ferguson, E. S., 171 W. 79th St., New York, N. Y.
Forest, H. W., 1515 Bernal Ave., Burlingame, Calif.
Frost, F. N., 611 E. Ninth, Port Angeles, Wash.
Green, J. E., 7509 Durocher, Montreal, Que., Canada
Hastings, C. L., 5124 W. Moncrieg, Denver, Colo.
Hills, E. G., 3001 W. Main St., Belleville, Ill.
Horn, L. H., 207 E. Ohio St., Chicago, Ill.
Hornby, R. H., 525 E. Lewis St., Fort Wayne, Ind.
Hunt, G., R. Gutierrez 4667, Buenos Aires, Argentina
Ingling, T. M., 300 N. Jackson St., Belleville, Ill.
Intiso, J. F., 1730 Fillmore St., New York, N. Y.
Jackson, J. E., RCA Victor Company, Ltd., Montreal, Que., Canada
Kelly, J. J., c/o Station CKGB, Timmins, Ont., Canada
Knights, S. F., 1435 St. Matthew St., Montreal, Que., Canada
Langhoff, R., 248 Steuben St., Pittsburgh, Pa.
Leonard, C. G., 13th School Squadron, Scott Field, Ill.
Lincoln, R. B., RCA Manufacturing Company, Inc., Harrison, N. J.
Litton, F., 1507 N. La Brea, Inglewood, Calif.
Marriott, B., 4031 Cartier St., Montreal, Que., Canada
Narayanan, P., Radio Electric Institute, Bombay 4, India
Peale, A. H., California-Arabian Standard Oil Company, Bahrein Island, Persian Gulf
Ragazzini, J. R., 515 W. 122nd St., New York, N. Y.
Respondek, A. M., Box 47, Belleville, Ind.
Rudnick, P., RCA Manufacturing Company, Inc., Camden, N. J.
Singh, G., Box 150, Delhi, India
Soper, L. M., 2334 Beverly Blvd., Los Angeles, Calif.
Striker, G. O., 3612 Franklin Blvd., Chicago, Ill.
Thomson, B., Jr., 733 Soldano Ave., Azusa, Calif.
Thukaram, Y. A., 56/A Big St., Triplicane, Madras, India
Van Campen, M., 632 Catherine Ave., Muskegon, Mich.
Vergason, D. L., Hygrade Sylvania Corporation, Emporium, Pa.
Waddington, J. K., "St. Mungo," 1st Beach, Clifton, Cape Town, South Africa
Wallace, R. B., 3 Wireless School, RCAF, Winnipeg, Manit., Canada
Webb, R. C., Elec. Eng. Bldg., Purdue University, Lafayette, Ind.
Weiler, G. W., 1001 Olive, Belleville, Ill.
Whitaker, G. C. F., c/o H. M. Naval Base, Singapore, Malaya
Williamson, L. C., 16509 Manor Ave., Detroit, Mich.
Wolf, H., 114-01-86th Ave., Richmond Hill, L. I., N. Y.

Contributors

S. J. Begun was born in the free city of Danzig on December 2, 1905. He received his Master's degree in electrical engineering in 1929 from the Institute of Technology in Berlin, and in 1933 he received the Doctor's degree. During 1928–29 he was with the Automatische Telephon, Berlin, where he designed circuits for automatic long-distance telephone systems; from 1929 to 1930 with Ferdinand Schuchhardt, working with magnetic, disk, and film sound-recording equipment; from 1930 to 1932 with the Echophon Maschinen working with development and sales of recording and dictating machines; from 1932 to 1935 with C. Lorenz, where he had charge of the Electro-Acoustic Laboratory and supervised production of technical specialties; from 1935 to 1937 as chief of the Development Laboratory of Guided Radio where he developed emergency communication systems for ships; from 1937 to 1938, associated with Acoustic Consultants, where he did consulting work in connection with room acoustics, public-address systems, and other acoustic problems. Since 1938 Dr. Begun has been with the Brush Development Company supervising the design and production of magnetic disk systems.



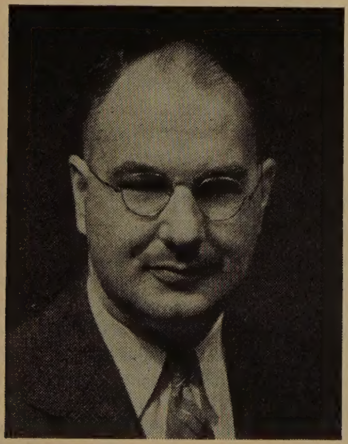
S. J. Begun

❖

Charles M. Burrill (A'24–M'30) studied electrical engineering at the University of Minnesota, graduating in 1923. He then went with the General Electrical Company. Following three years of general training in the General Electric advanced course in engineering, he joined the radio engineering department, and in 1927 was placed in charge of tuned-radio-frequency receiver development. Since 1930 Mr. Burrill has been with RCA at Camden, N. J., with the exception of a year and a half in 1931–1932 spent with the Rogers-Majestic Corporation of Toronto, Canada, in charge of research. Since returning to Camden he has been engaged in general research, first in sound recording, and more recently in interference and noise suppression and in ultra-short-wave propagation. He is a member of Tau Beta Pi, Eta Kappa Nu, Sigma Xi, and the Franklin Institute.

❖

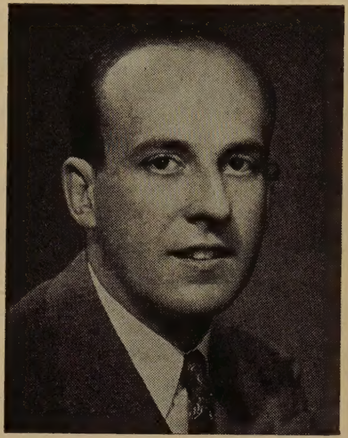
Alan M. Glover (A'39) was born in Rochester, New York, on September 26, 1909. He received the B.A. degree in 1930 from the University of Rochester; the M.A. degree in 1932, and the Ph.D. degree in Physics in 1935, also at the University of Rochester. In 1935 and 1936, he was on the staff of the Institute of Paper Chemistry, Lawrence College, Appleton, Wisconsin. Since 1936, he has been with the research and engineering department of the RCA Manufacturing Company, engaged in the development of phototubes and other special tube activities of the company.



Charles M. Burrill

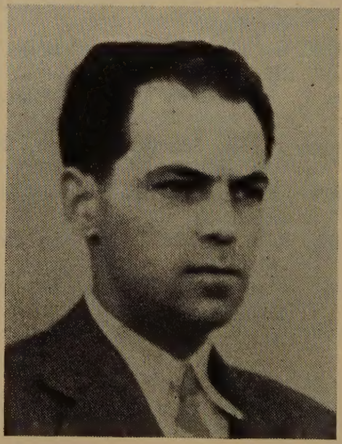
❖

Lloyd M. Hershey (M'39) was born on March 24, 1908, at Lafayette, Indiana. He received the B.S. degree in electrical engineering from Purdue University in 1930. From 1930 to 1932 he was employed by the Bell Telephone Laboratories. Mr. Hershey did production and design work for several different radio manufacturers in New York City between 1933 and 1936. Since 1936 he has been an engineer in the New York laboratory of the Hazeltine Service Corporation.



Stechbardt Studios

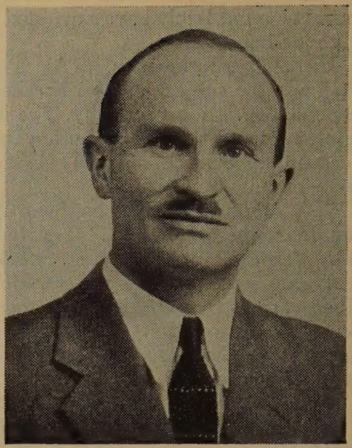
Alan M. Glover



Lloyd M. Hershey

❖

Ronold King (A'30) was born on September 19, 1905, at Williamstown, Massachusetts. He received the B.A. degree in 1927 and the M.S. degree in 1929 from the University of Rochester and the Ph.D. degree from the University of Wisconsin in 1932. He was an American-German exchange student at Munich from 1928 to 1929; a White Fellow in physics at Cornell University from 1929 to 1930; a Fellow in electrical engineering at the University of Wisconsin from 1930 to 1932. He continued at Wisconsin as a research assistant from 1932 to 1934. From 1934 to 1936 he was an instructor in physics at Lafayette College, serving as an assistant professor in 1937. During 1937 and 1938 Dr. King was a Guggenheim Fellow at Berlin. In 1938 he became instructor in physics and communication engineering at Harvard University, advancing to assistant professor in 1939.



Ronold King

❖

Harold Alden Wheeler (A'27, M'28, F'35) was born at St. Paul, Minnesota, on May 10, 1903. He received his B.S. degree in physics at George Washington University in 1925. From 1925 to 1928 he took postgraduate work in the physics department of Johns Hopkins University and was a lecturer there from 1926 to 1927. From 1921 to 1922 he was an assistant in the Radio Section of the National Bureau of Standards. In 1923 Mr. Wheeler became an assistant to Professor Hazeltine and in 1924 he entered the Hazeltine Corporation and the Hazeltine Service Corporation as an engineer. He is a member of Sigma Xi.

❖

For a biographical sketch of C. K. Jen, see the Proceedings for June, 1941.

❖

H. A. Wheeler